Proteomic analysis by mass spectrometry

from basic to application

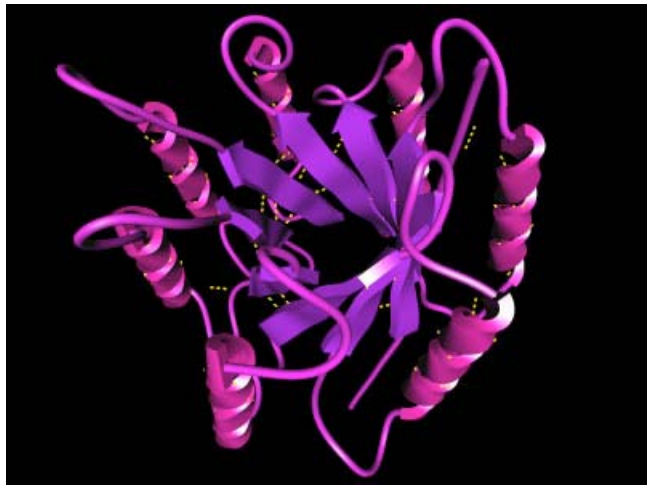
Zuzana Demianova
MS group
IOCB AS CR v.v.i

Outline

- Brief introduction to proteomics ideology
- Combination protein and MS
- MS data interpretation
- Applications

Protein, Proteome, Proteomics

Protein is a complex organic molecule that contain carbon, hydrogen, oxygen and nitrogen, is composed of one or more chains of amino acids.



Proteins are fundamental components of all living cells are necessary for the proper functioning of an organism.

Proteome is the set of expressed proteins at a given time under defined condition.

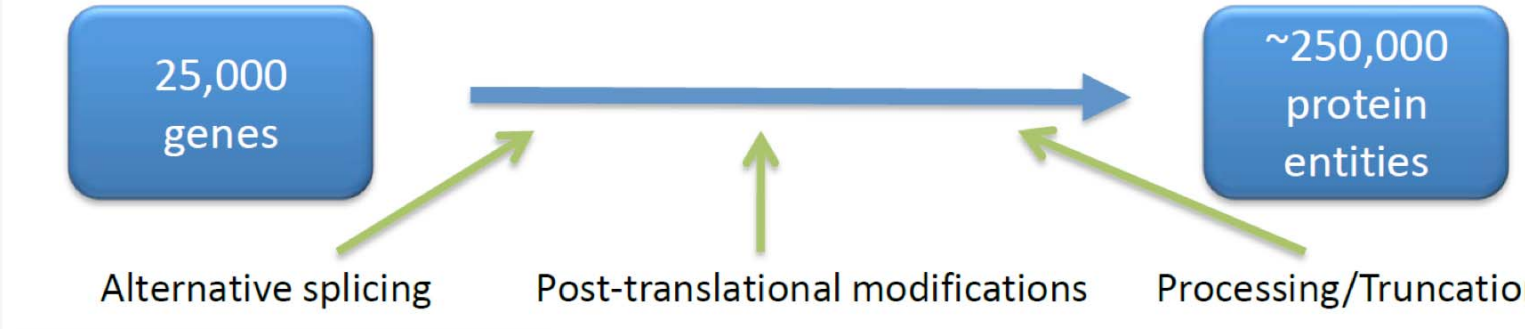
Proteomics is field of study encompassing the identification and quantification of proteins, and the effect of their modifications, interactions, activities, and function, during disease states, and treatment.

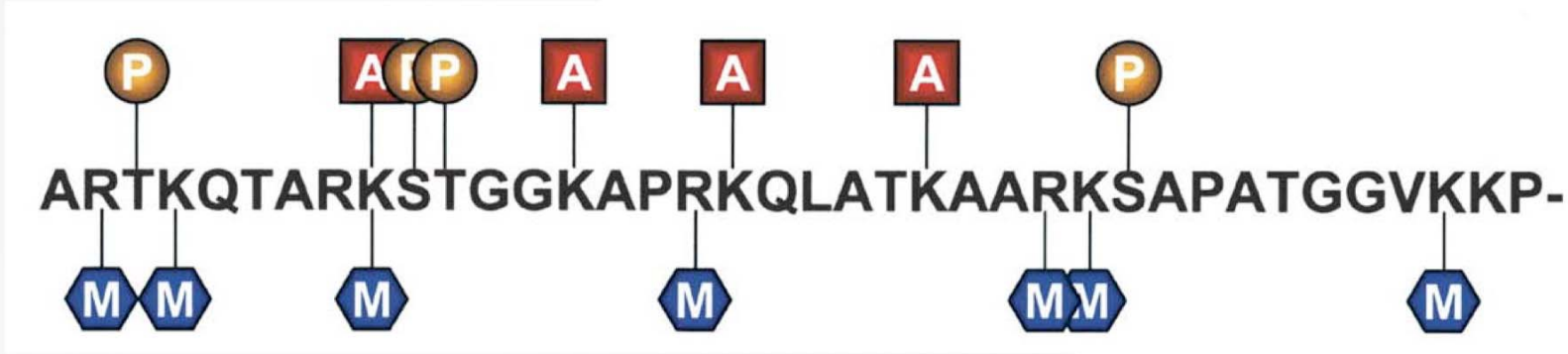
Charting the protein composition of a cell/organism, and its change as a response to internal/external cues.

Challenges in proteomics

- **Complexity** of the proteome
- **Dynamic range** of protein abundance

Complexity of the proteome

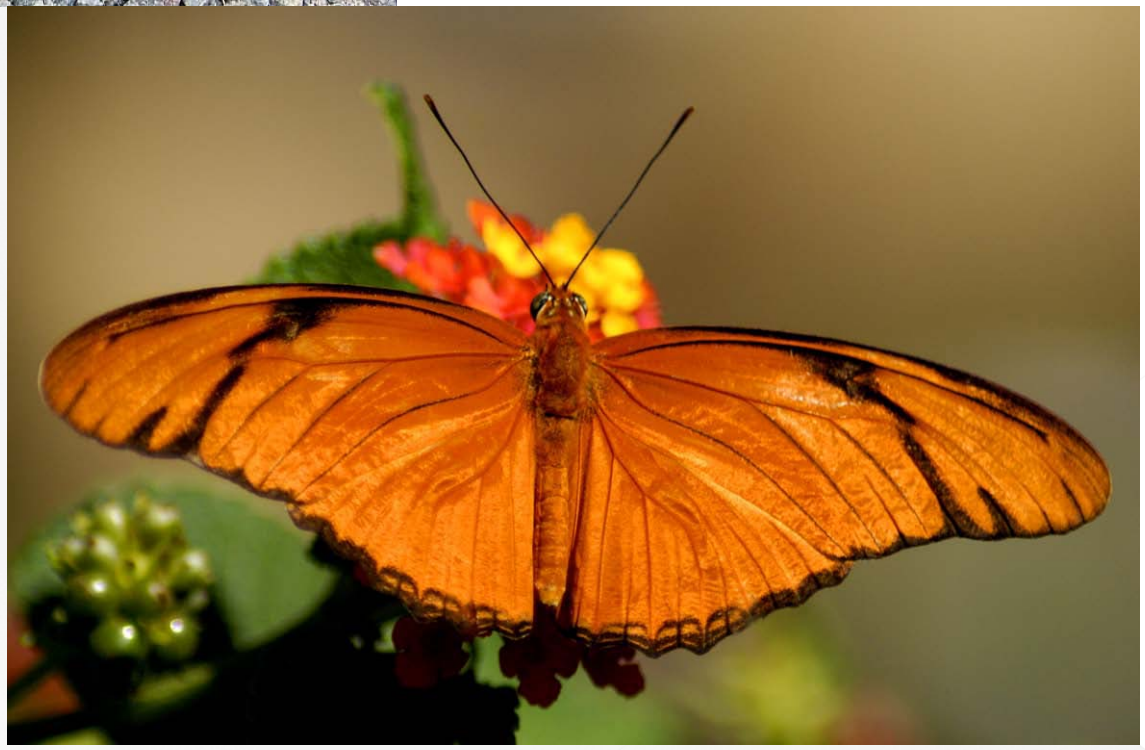




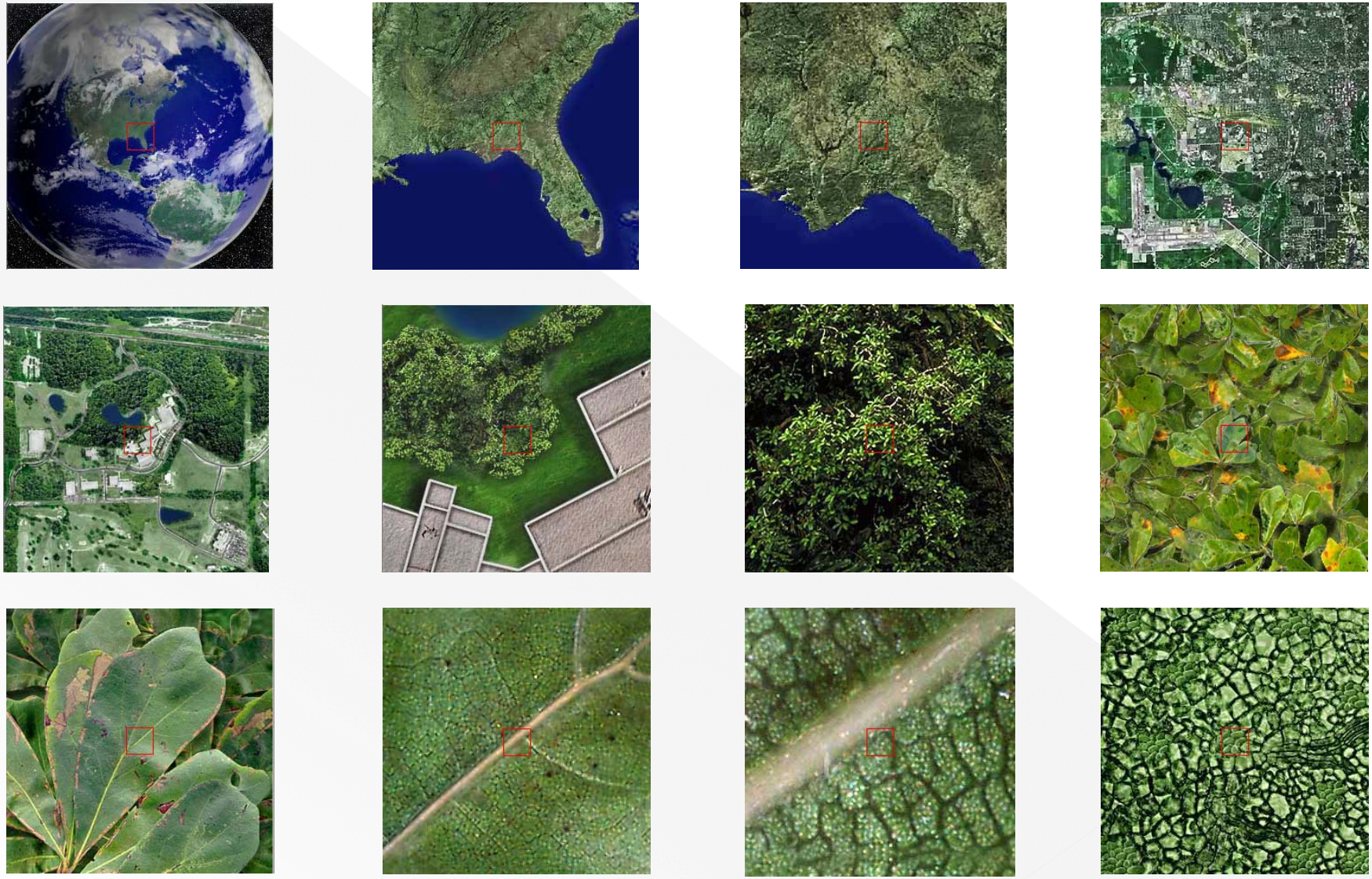
15 modifications on the N-terminal tail of histone 3

2^{15} = 32,000 combinations (at least in theory)

This alone exceeds the number of genes in the genome



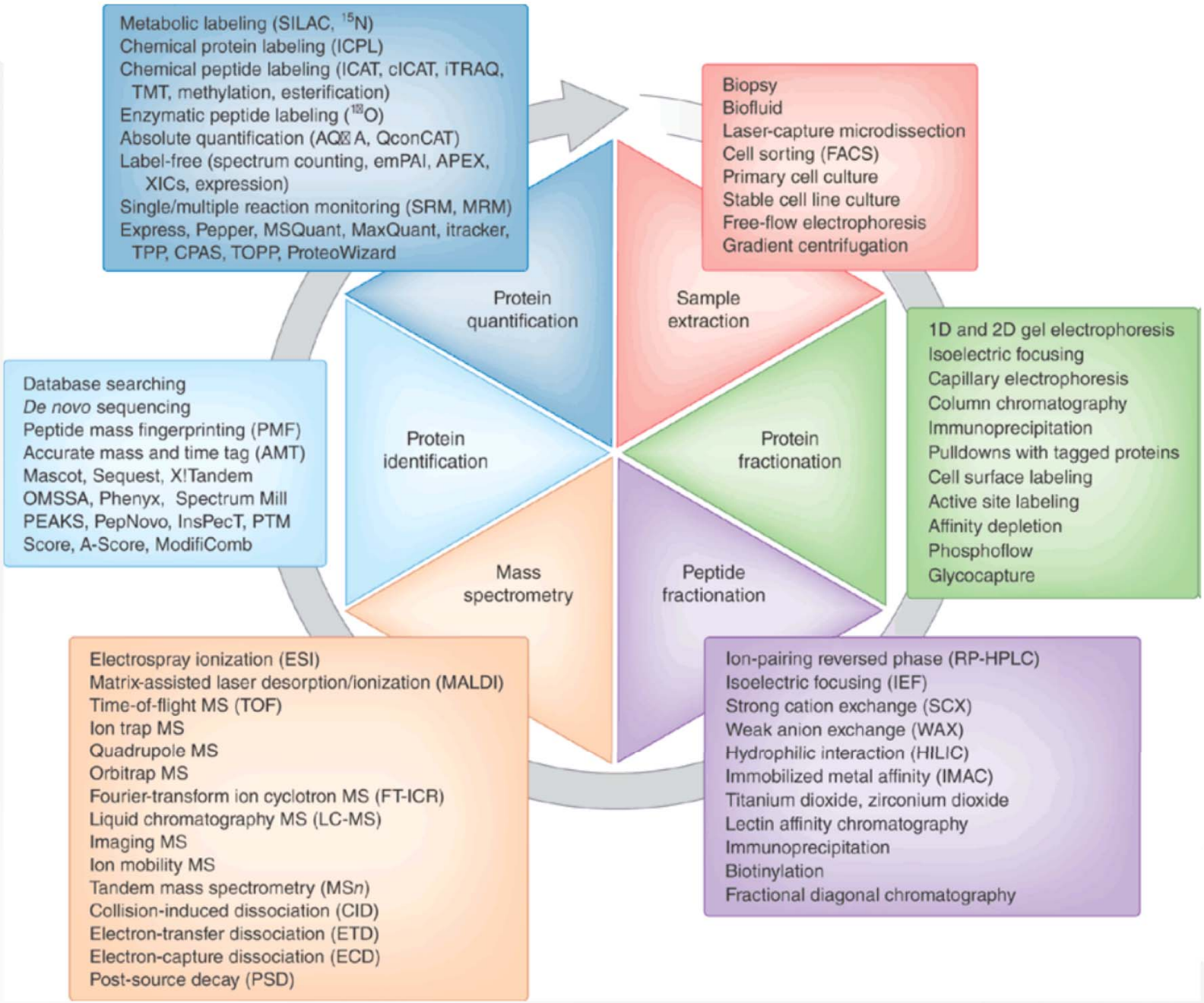
Dynamic range of serum proteins: 12 orders of magnitude



‘I want a camera that can capture continents and single cells’

Albumin

Interleukin 6

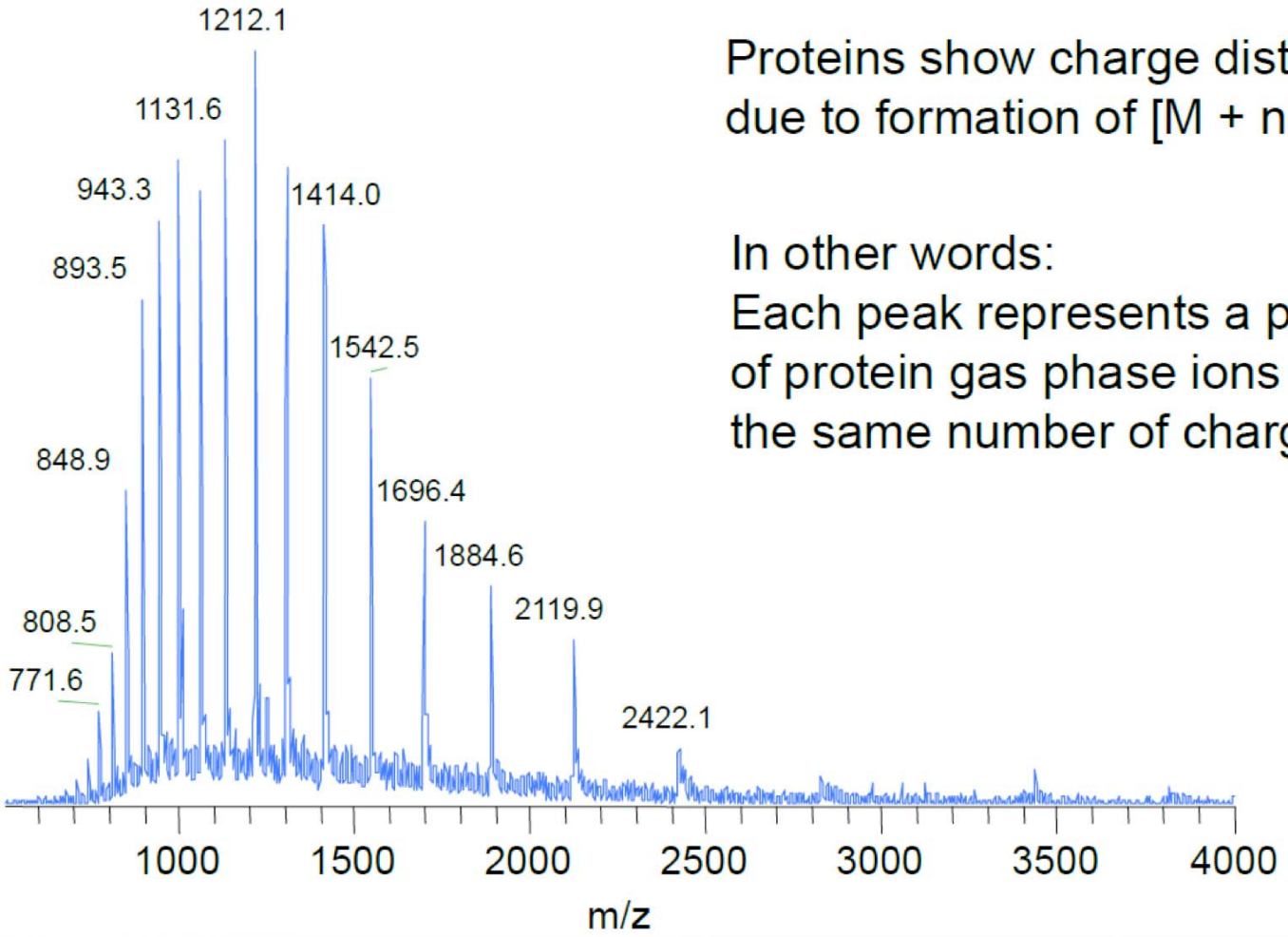


Mallick& Kuster, Nature Biotechnology 28, 695–709 (2010)

Protein Identification

- How can we identify a protein by MS?
- Why do we need to digest protein?
- How can we identify 1000 proteins by MS?
- How do we know they are correct?
- How do we quantify identified protein by MS?

ESI-MS of denatured myoglobin

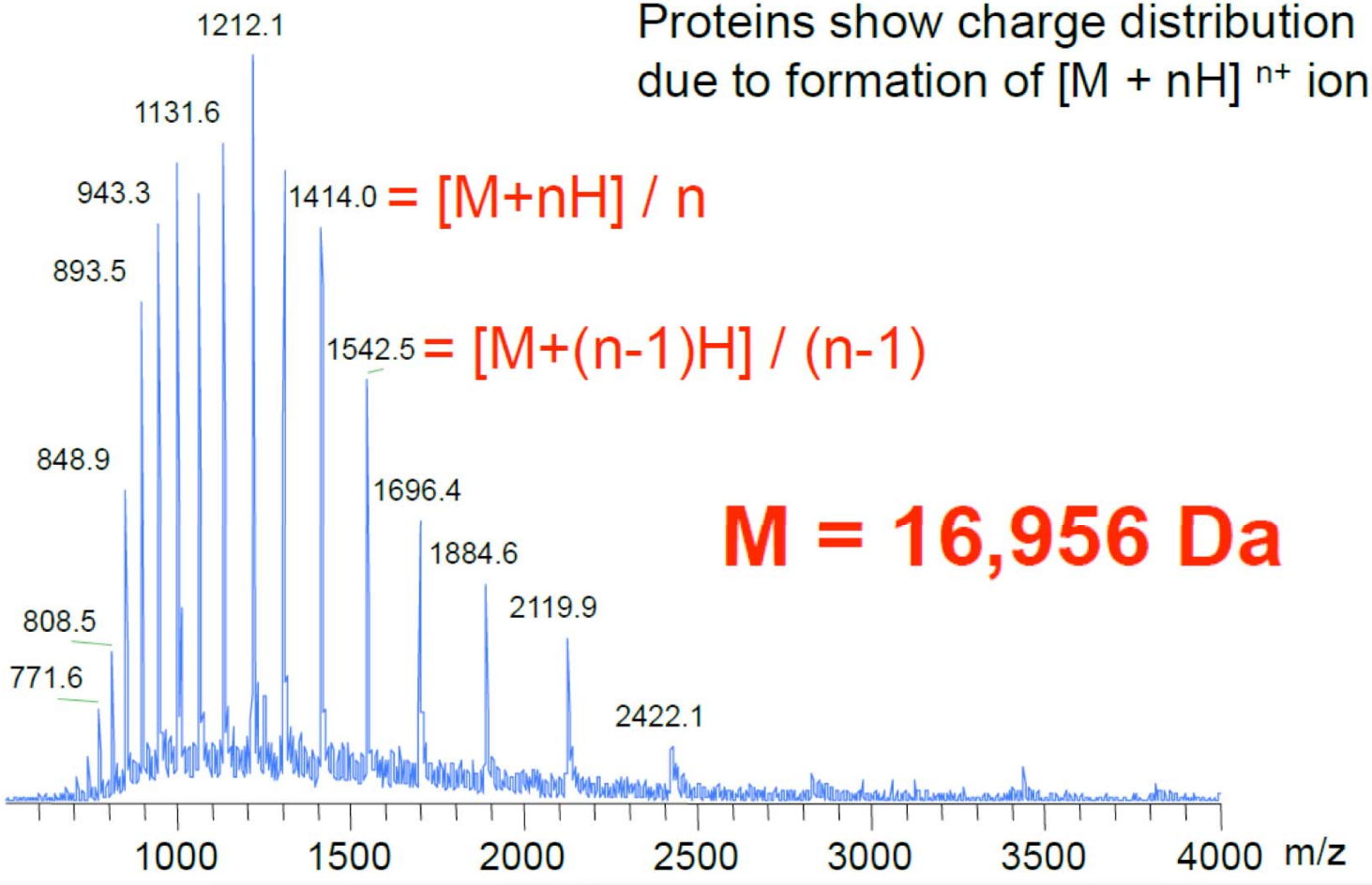


Proteins show charge distribution due to formation of $[M + nH]^{n+}$ ions

In other words:
Each peak represents a population of protein gas phase ions carrying the same number of charges

ESI-MS of denatured myoglobin

How to calculate its molecular weight?



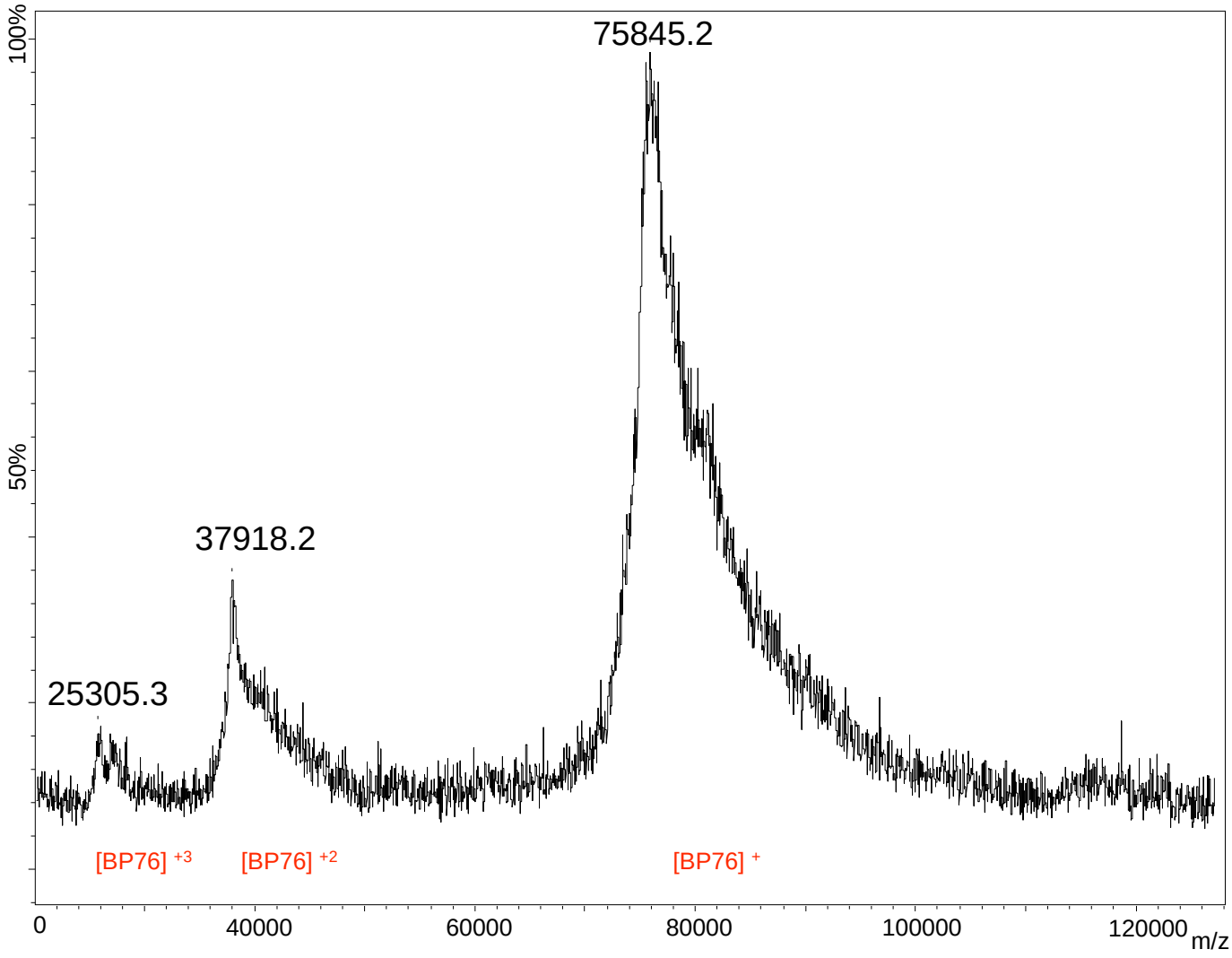
Proteins show charge distribution due to formation of $[M + nH]^{n+}$ ions

$= [M+nH] / n$

$= [M+(n-1)H] / (n-1)$

M = 16,956 Da

MALDI-TOF-MS of BP76

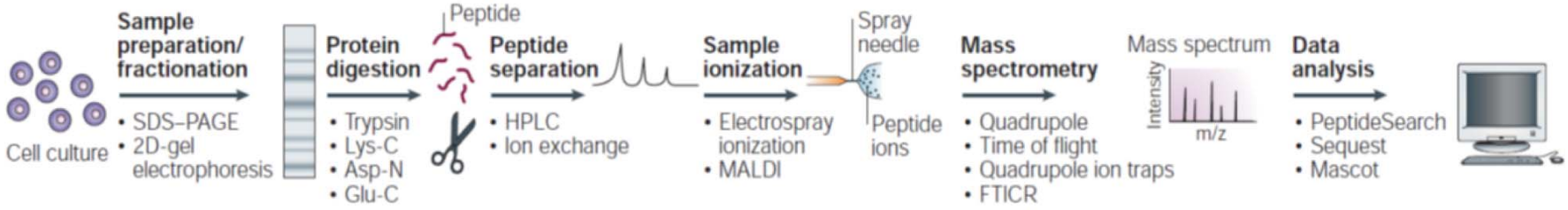


Protein Identification

- How can we identify a protein by MS?
- Why do we need to digest protein?
- How can we identify 1000 proteins by MS?
- How do we know they are correct?
- How do we quantify identified protein by MS?

Protein Identification by MS

Bottom Up approach

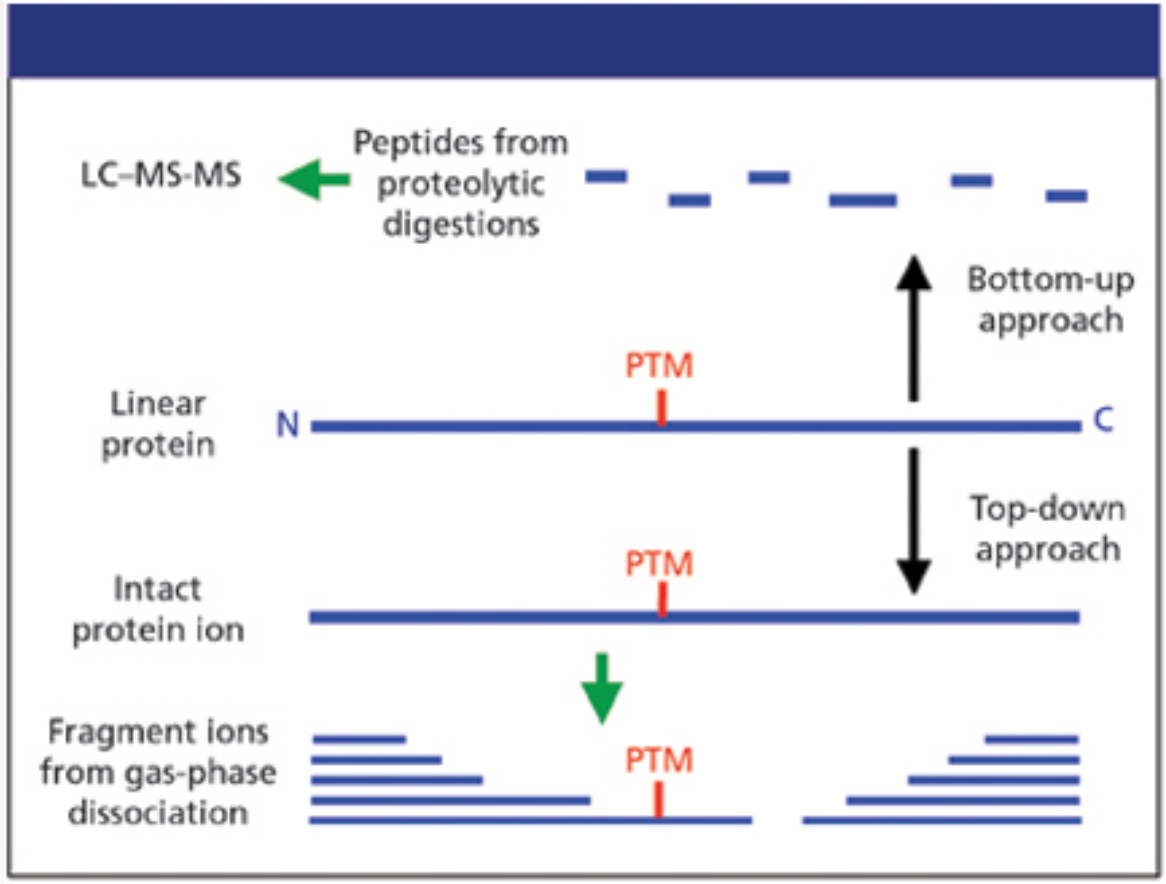


- Gel based methods (SDS-PAGE, 2DE)
- Gel-free based methods (1-D LC, MuDPIT, Shotgun proteomics)
- Protein Identification and differential proteomics
- CID or ECD

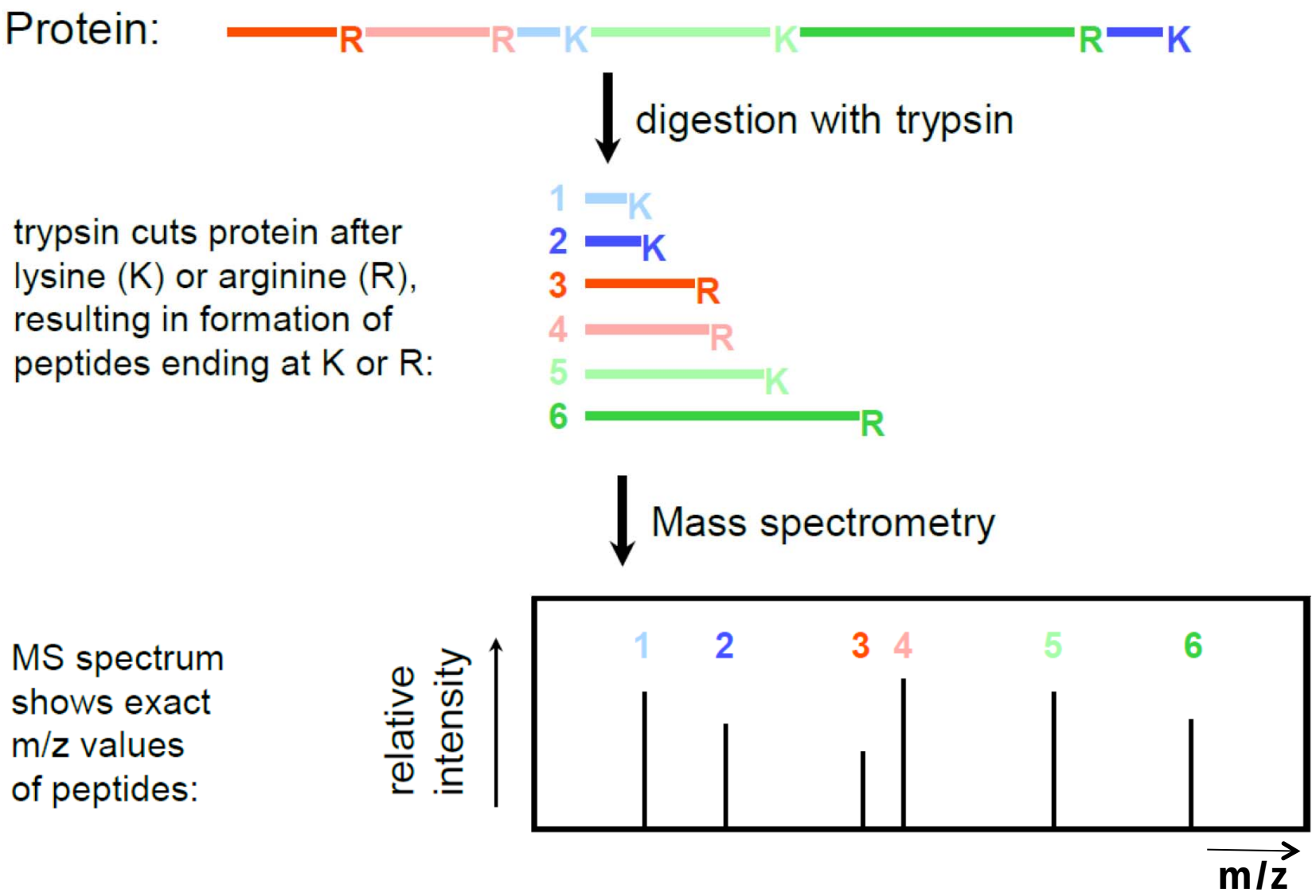
Top Down approach

- Separation of simple mixtures of proteins
- Potential acces to complete protein sequence
- Ability to localize and characterize PTMs
- ECD

Bottom Up proteomics vs. Top Down proteomics



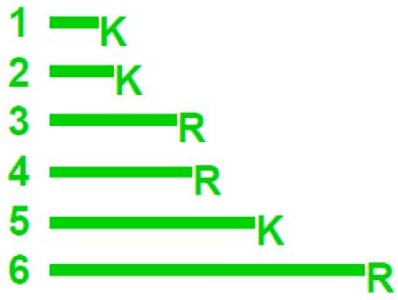
Peptide Mass FingerPrinting



Peptide Mass FingerPrinting

1.

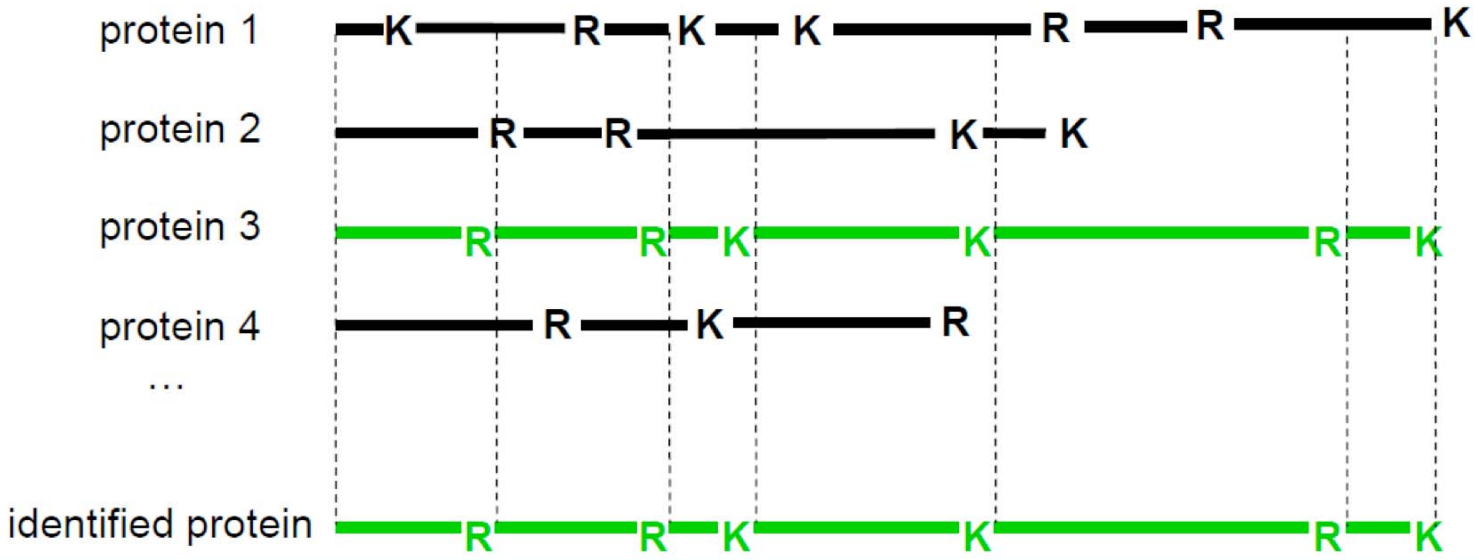
peptide map of target protein:



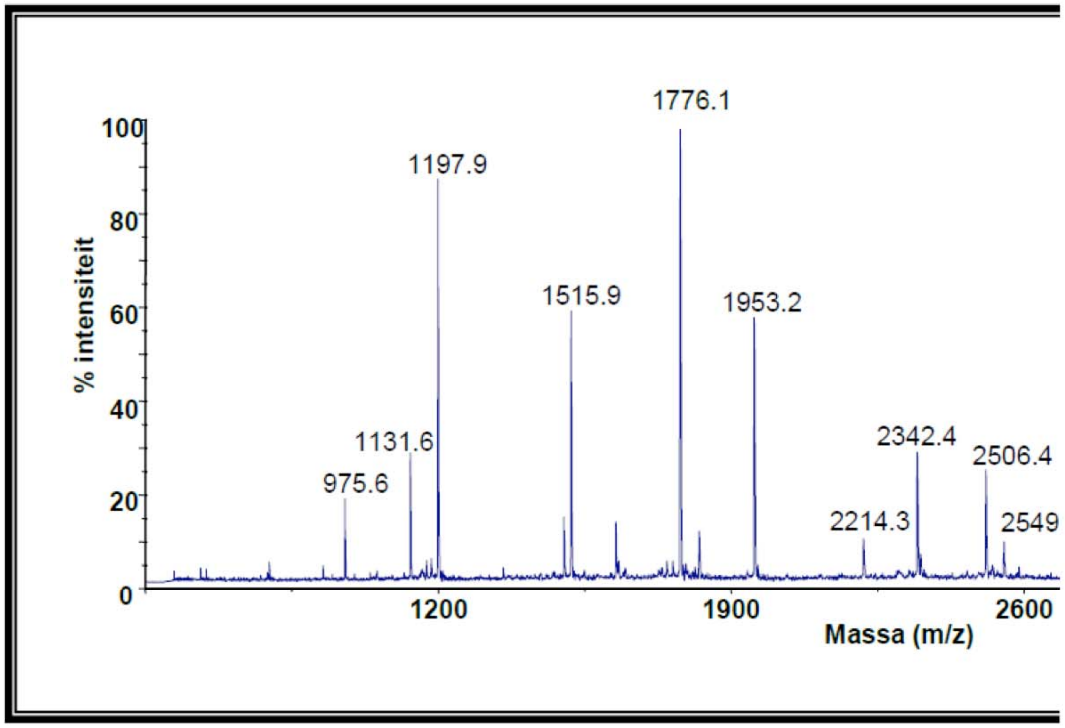
MALDI-TOF mass spectrum:

2.

Database search: matching experimental masses with *in silico* digest.

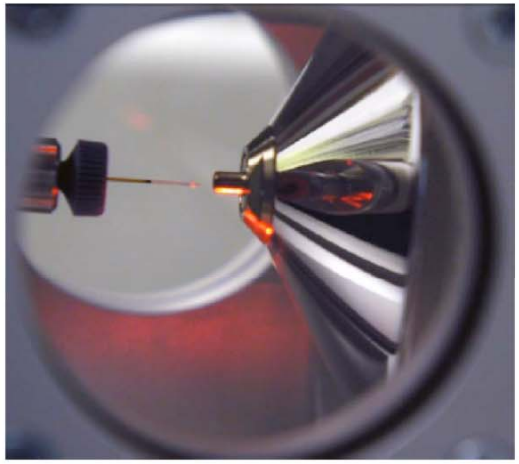


Peptide Mass FingerPrinting



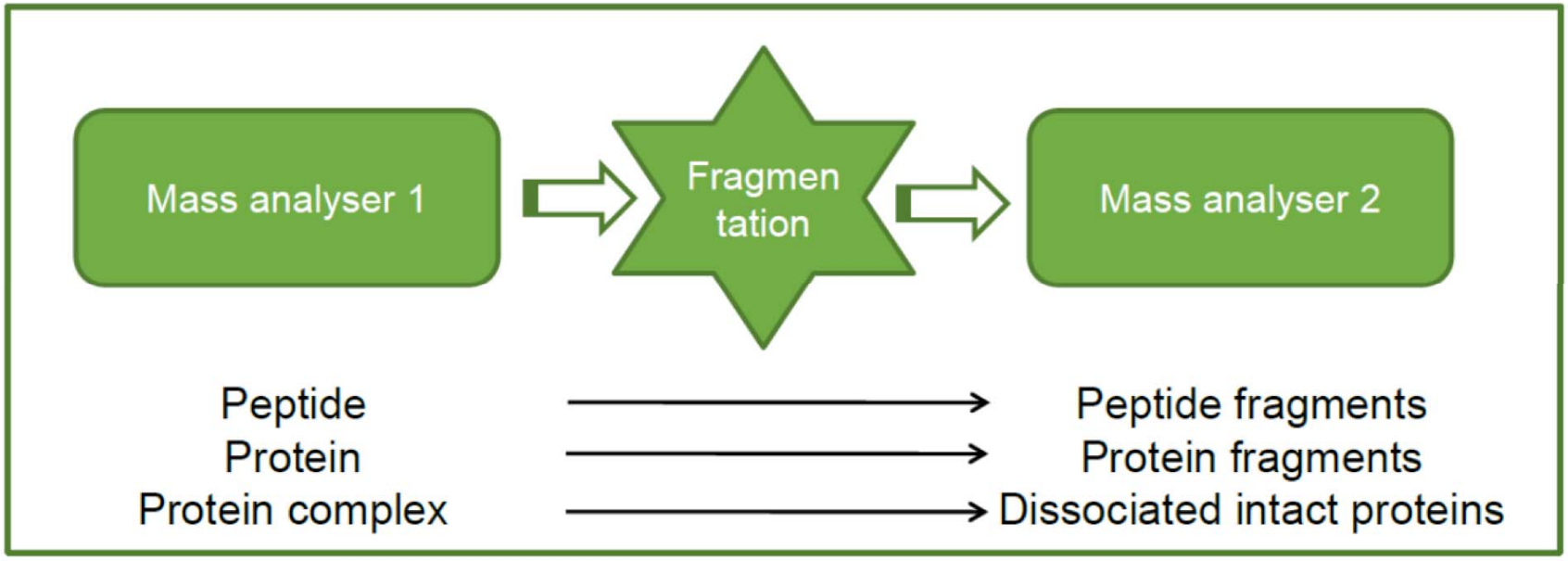
PMF is not amenable for the identification of protein mixtures

Peptide Fragmentation (MS/MS) and coupling to LC (LC-MS/MS)

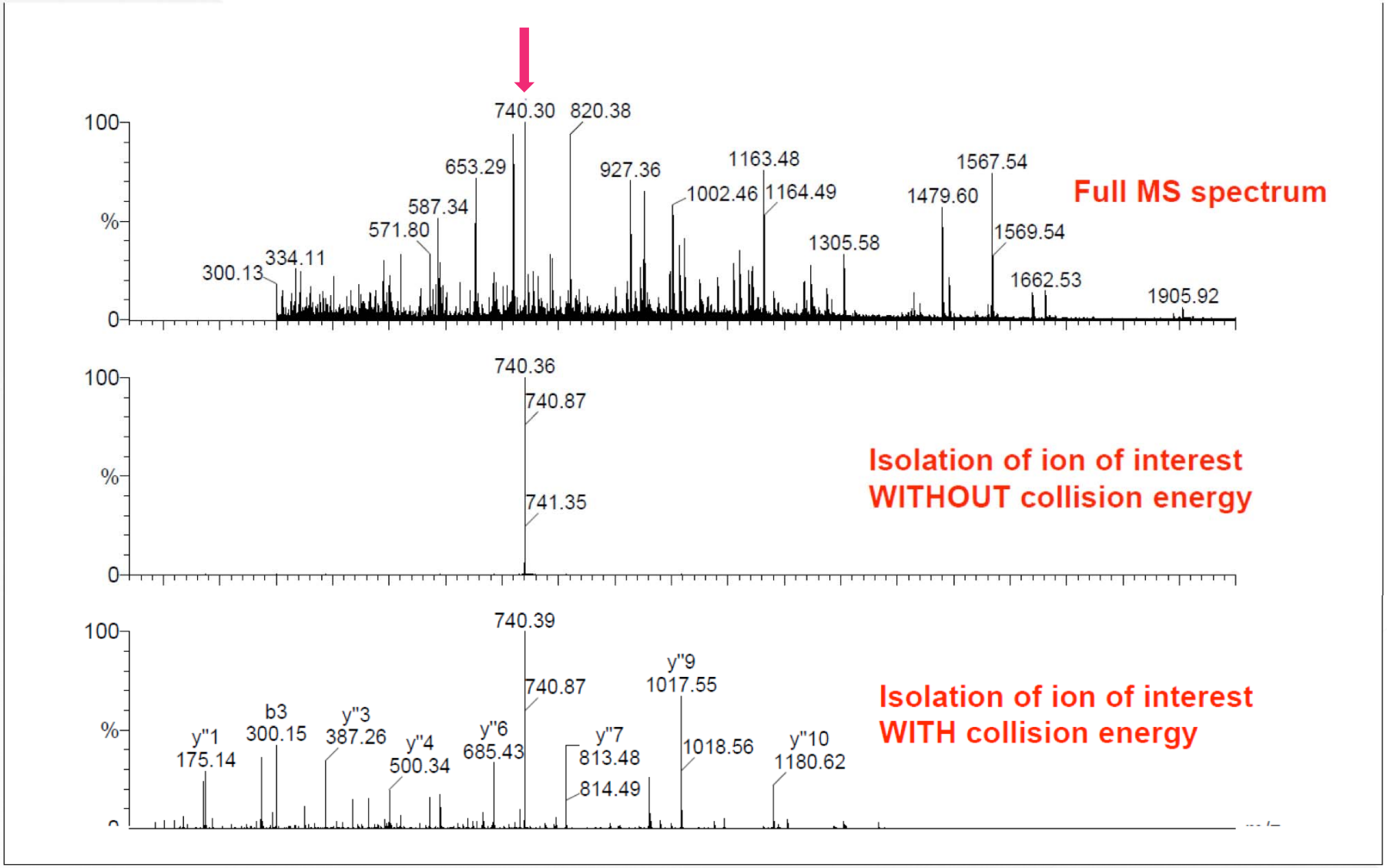


Tandem Mass Spectrometry

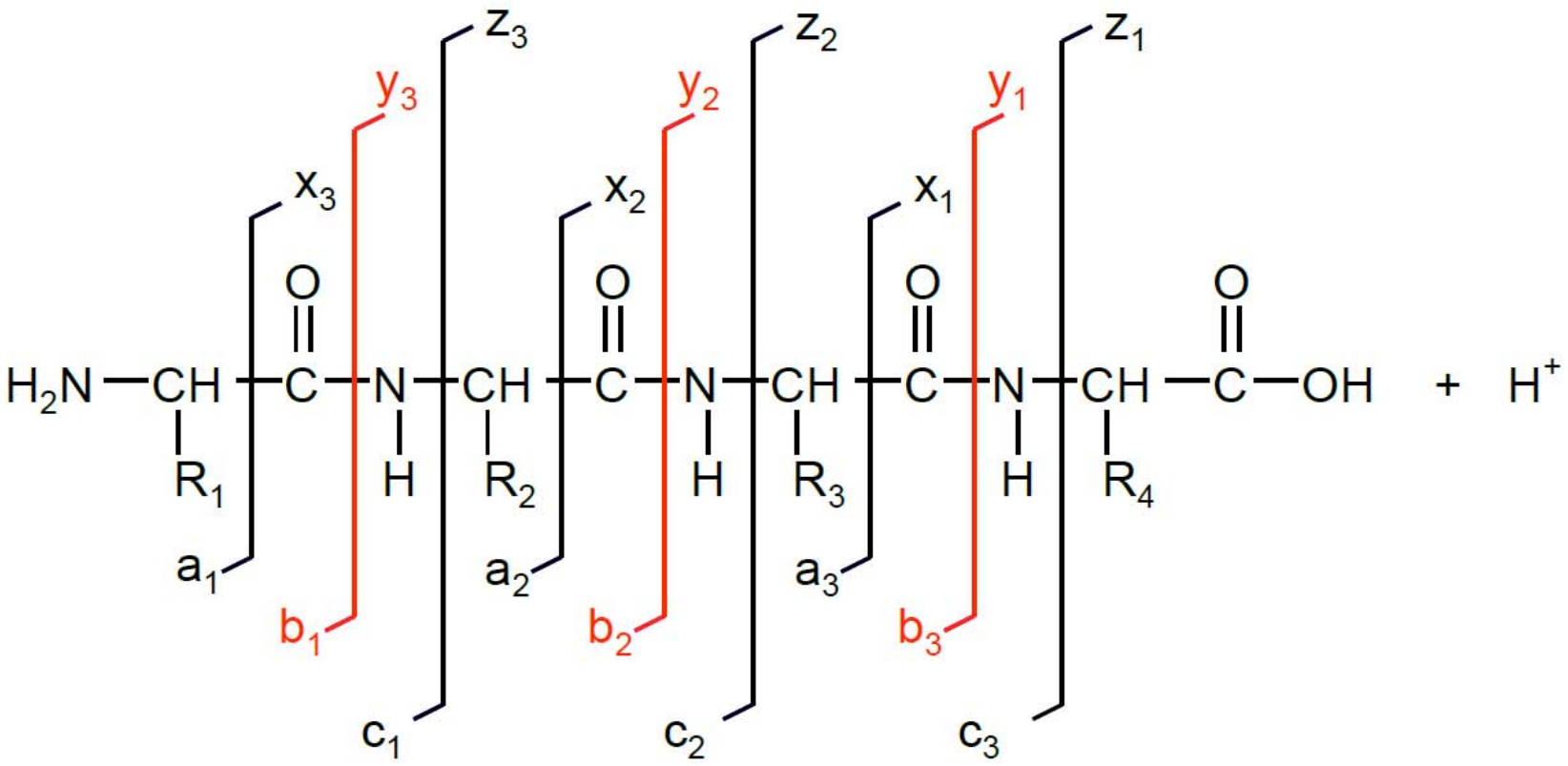
- Purpose is to fragment ions from parent ion to provide structural information about a molecule
- Uses two or more mass analyzers/filters separated by a gass-filled collision cell (Nitrogen, Argon or Xenon)



Tandem Mass Spectrometry



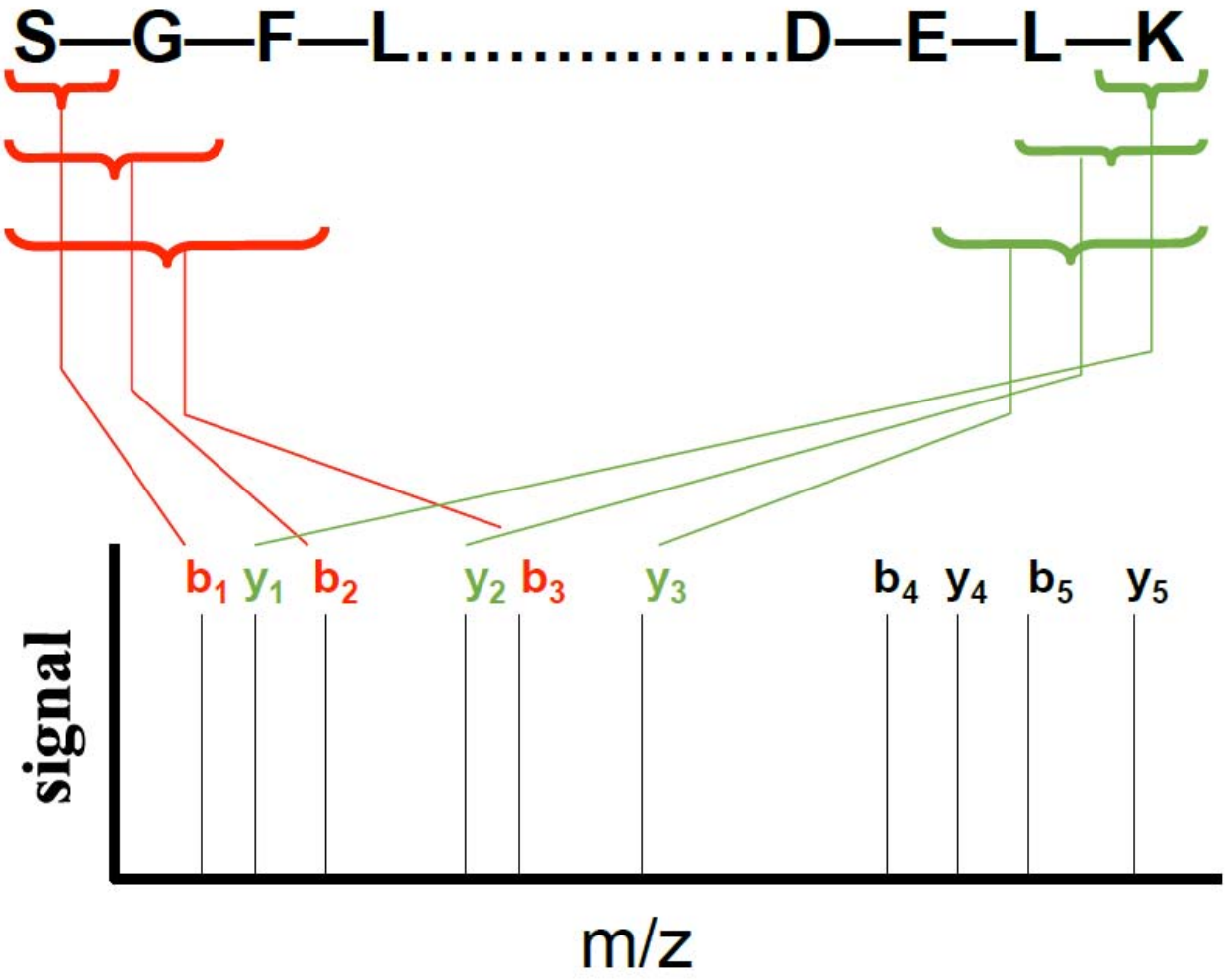
Peptide Fragmentation



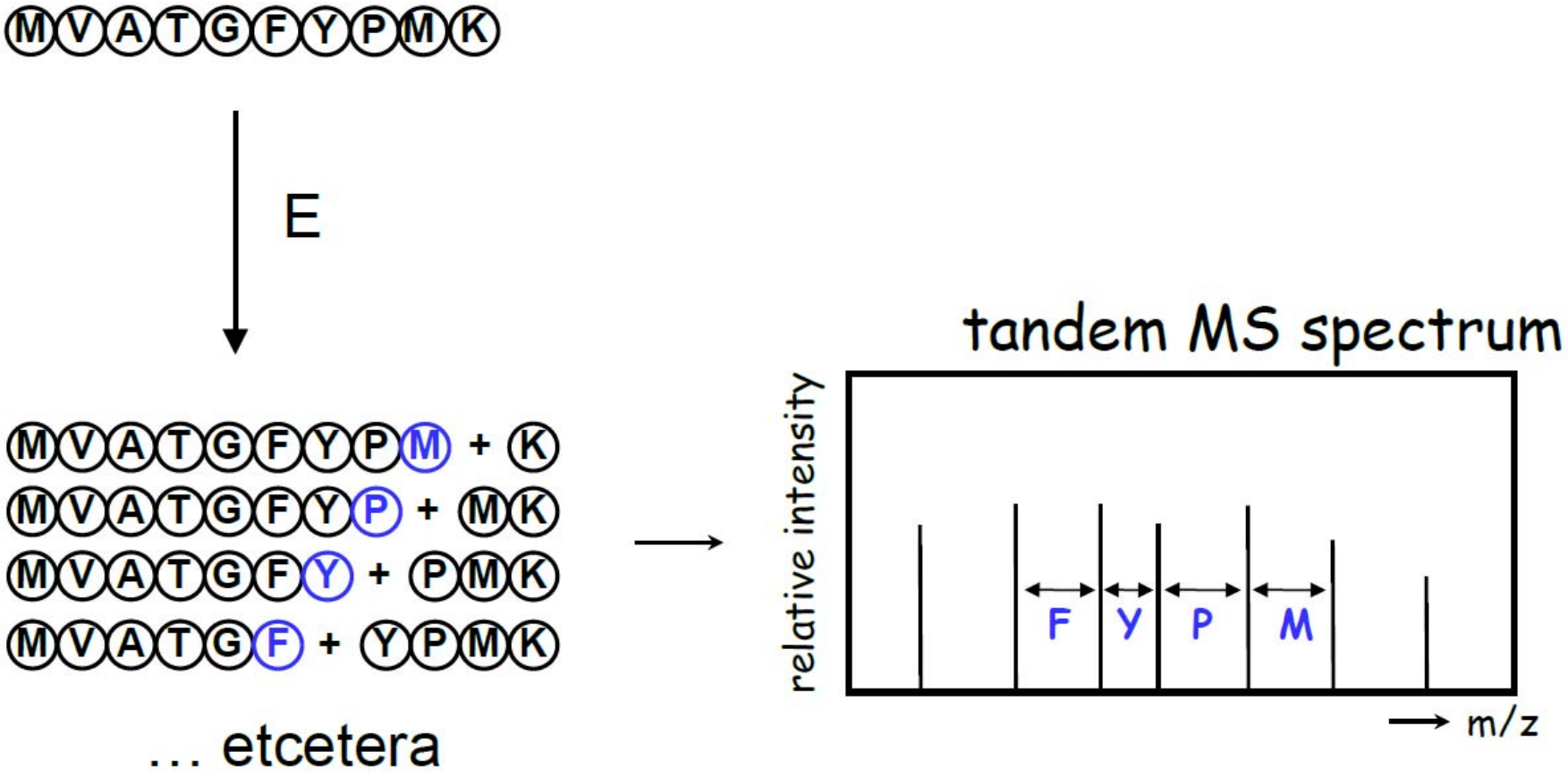
- Low energy fragmentation: break weakest bond = peptide bond
- Yields primarily b and y ions

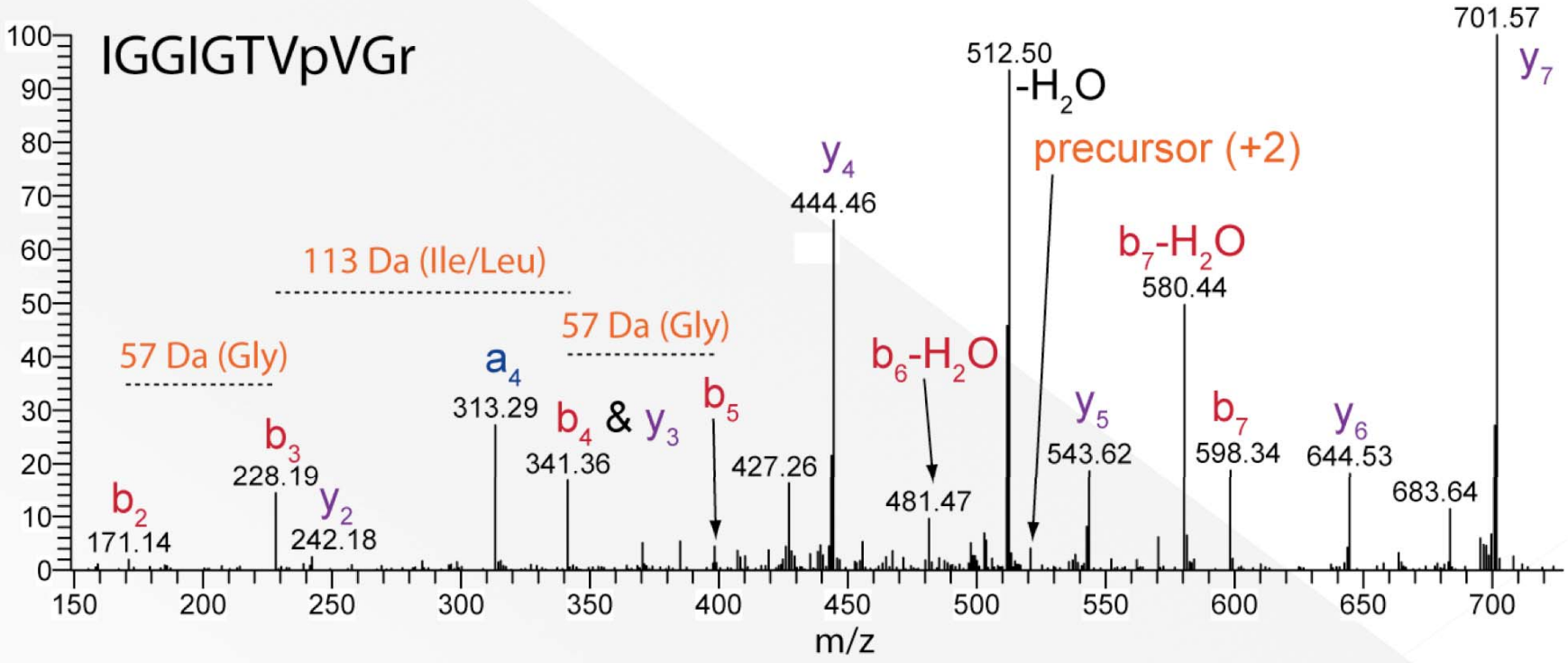
Peptide Fragmentation

In reality, you will see a combination of b- and y-ions



Peptide Fragmentation



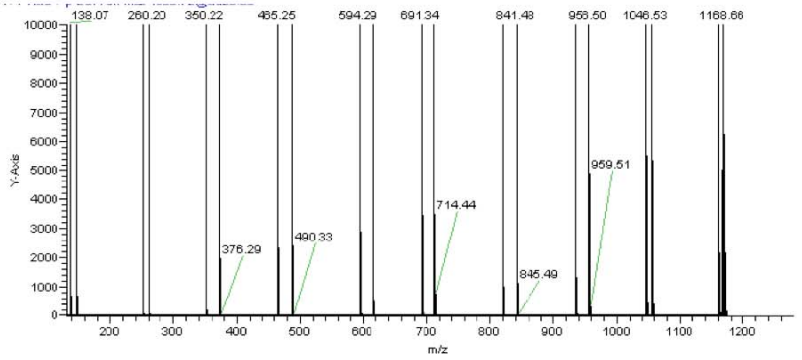

IGGIGTVpVGr
100
90
80
70
60
50
40
30
20
10
0
150
200
250
300
350
400
450
500
550
600
650
700
m/z
701.57
y7
512.50
-H2O
precursor (+2)
y4
444.46
b7-H2O
580.44
113 Da (Ile/Leu)
57 Da (Gly)
57 Da (Gly)
a4
313.29
b6-H2O
b5
b4 & y3
341.36
y5
543.62
b7
598.34
y6
644.53
683.64
b3
228.19
427.26
481.47
b2
171.14
y2
242.18

Peptide fragmentation is predictable: basis for automated interpretation

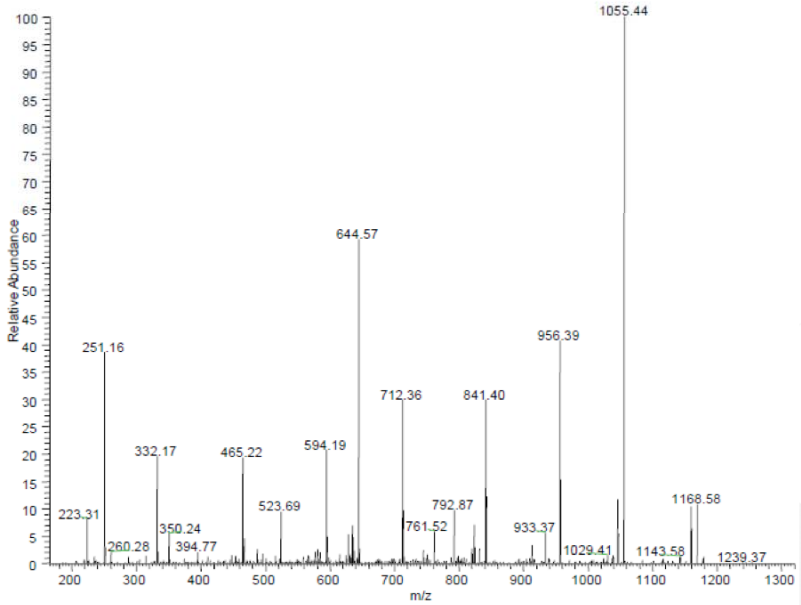
Peptide sequence:
HLVDEPQNLIK

No		B	Y
1	H	138.06	-
2	L	251.15	1168.65
3	V	350.21	1055.57
4	D	465.24	956.50
5	E	594.28	841.47
6	P	691.34	712.43
7	Q	819.39	615.38
8	N	933.44	487.32
9	L	1046.52	373.28
10	I	1159.61	260.19
11	K	-	147.11

theoretical



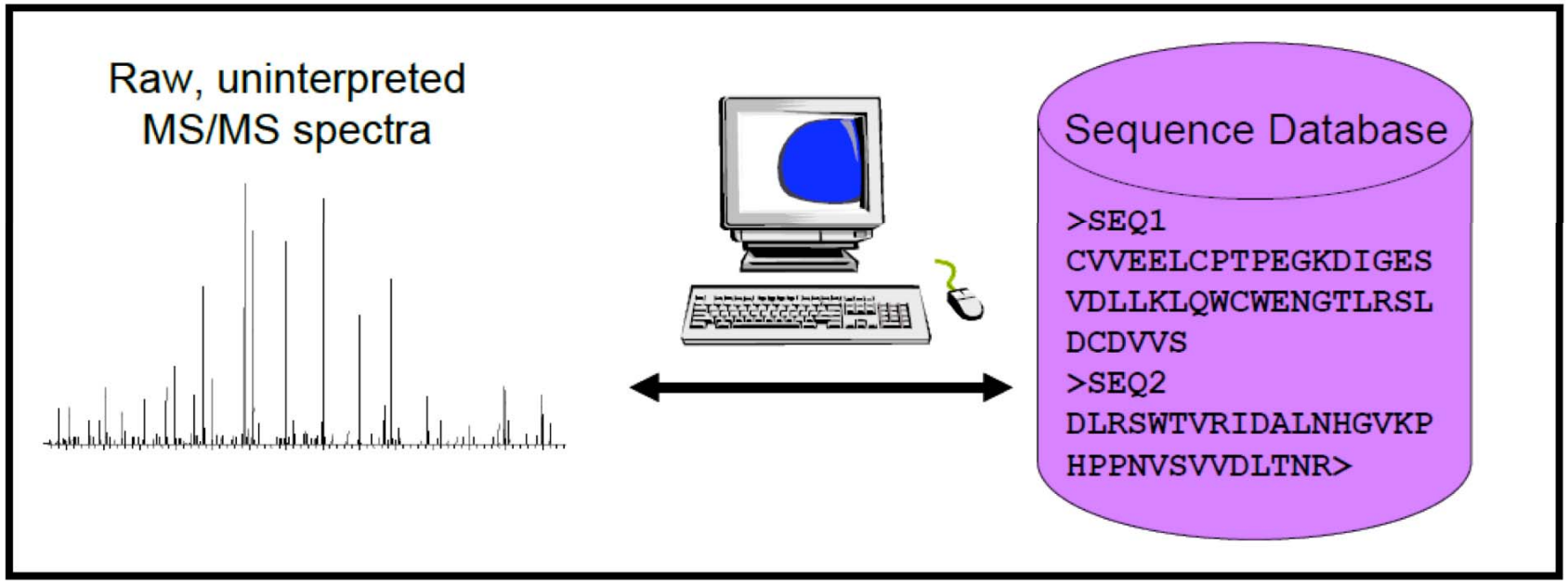
experimental



Bioinformatics and peptide identifications

Searching a protein database

- Matching fragmantation spectra to masses that can be formed from tryptic peptides in a particulat protein database (e.g. Swiss-Prot, Human Proteome, Ensemble...)
- Examples: Sequest, Macot, X!tandem, Phenyx, etc.

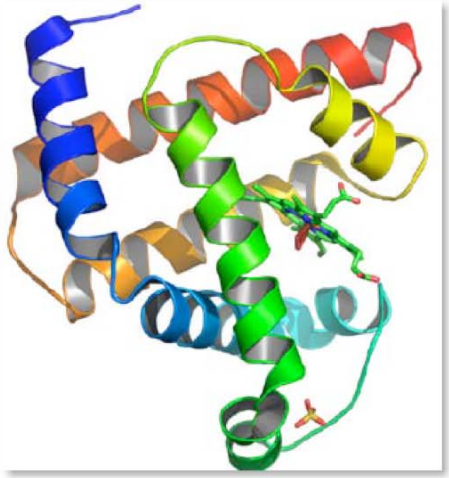


Protein Identification

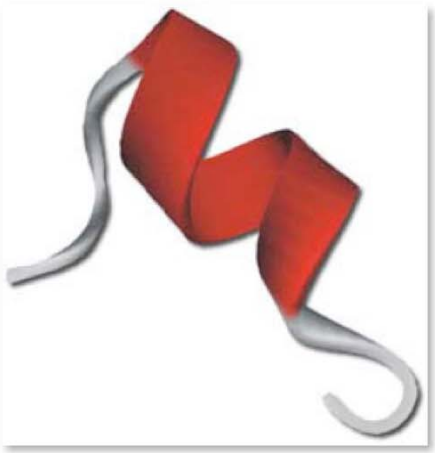
- How can we identify a protein by MS?
- Why do we need to digest protein?
- How can we identify 1000 proteins by MS?
- How do we know they are correct?
- How do we quantify identified protein by MS?

Protein Complexity

25,000
genes



100,000
proteins



1,000,000
peptides

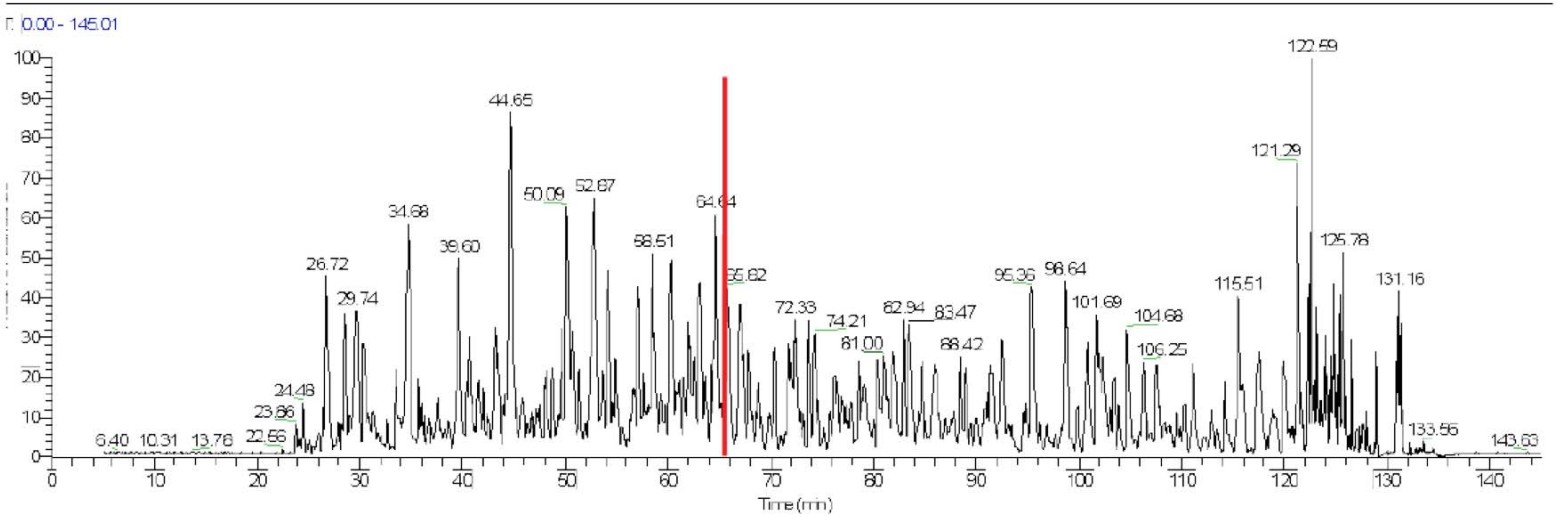
Solution to proteome complexity:

separation technique to deliver peptides at the MS one by one

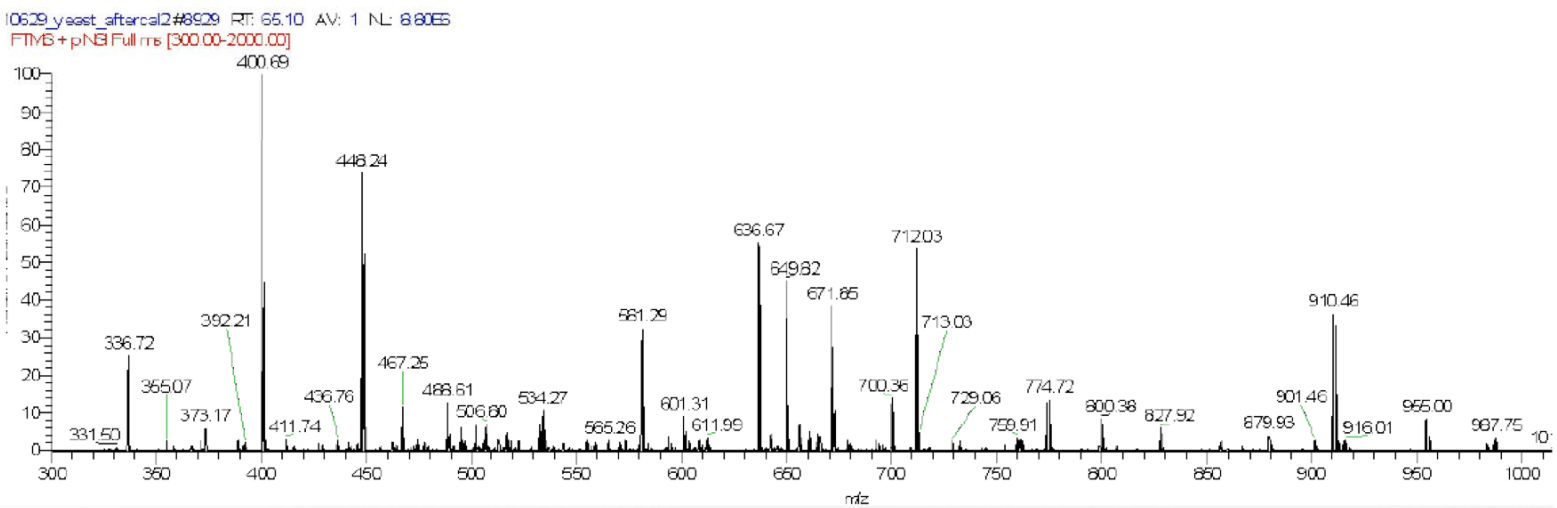
- Mass spectrometer filled with smaller diversity of peptides
- Less ion suppression (in electrospray)
- With proper separation, lower abundant peptides might be separated from higher abundant ones

- Often used methods: reverse-phase chromatography in combination with ion exchange chromatography

ion chromatogram



Spectrum at 65.1 min

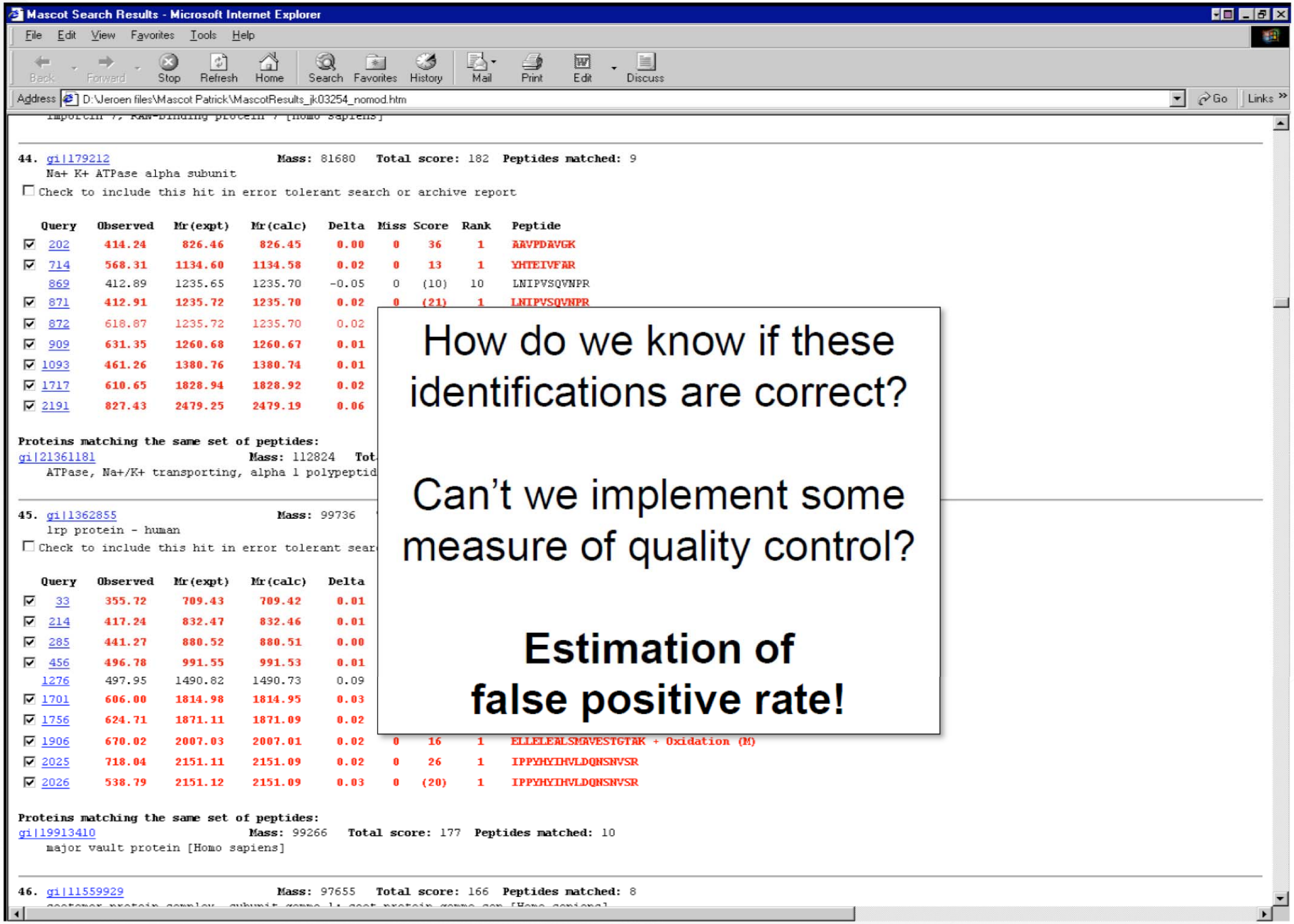


From MS to MS/MS: topN fragmentation

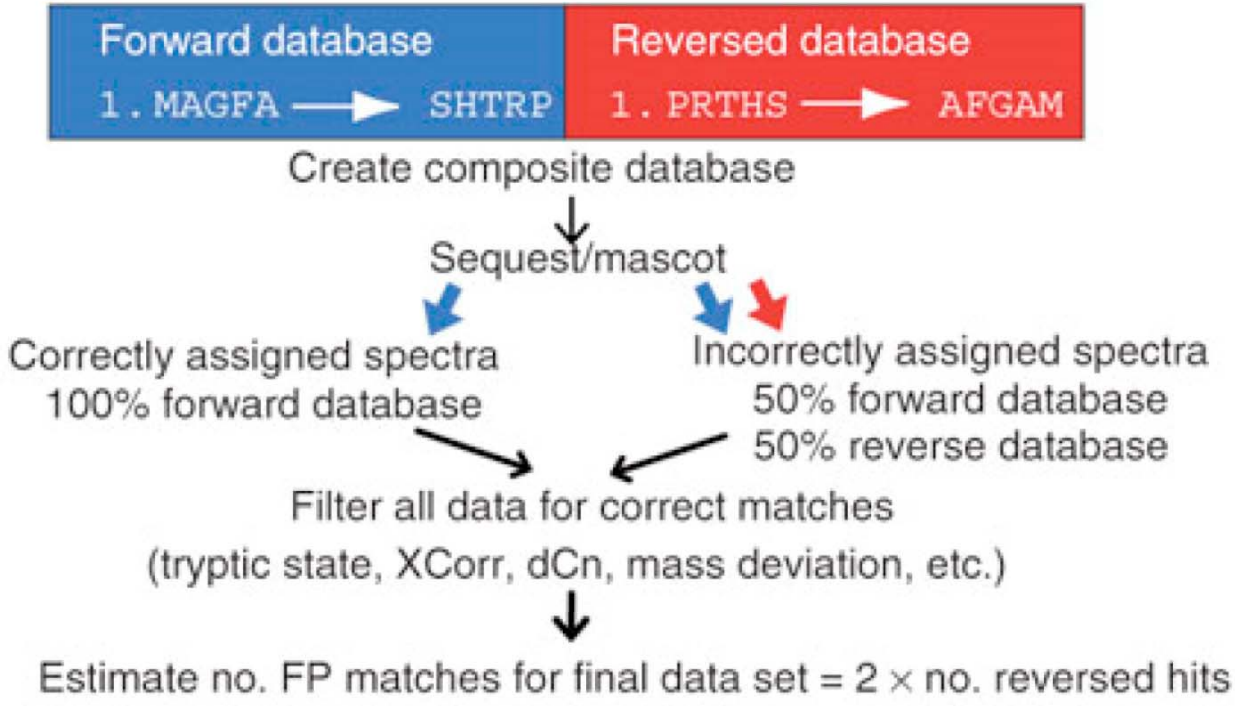
- Cycling through 1 x MS and N x MS/MS
- In LTQ-Orbitrap and Triple-TOF: MS and MS/MS in parallel
- Aiming for time-efficiency
- Dynamic exclusion: fragment every peptide only once (ideally)

Protein Identification

- How can we identify a protein by MS?
- Why do we need to digest protein?
- How can we identify 1000 proteins by MS?
- How do we know they are correct?
- How do we quantify identified protein by MS?


Mascot Search Results - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Forward Stop Refresh Home Search Favorites History Mail Print Edit Discuss
Address D:\Jeroen files\Mascot Patrick\MascotResults_jk03254_nomod.htm
Go Links
44. gi|179212 Mass: 81680 Total score: 182 Peptides matched: 9
Na+ K+ ATPase alpha subunit
Check to include this hit in error tolerant search or archive report
Query Observed Mr(expt) Mr(calc) Delta Miss Score Rank Peptide
202 414.24 826.46 826.45 0.00 0 36 1 AAVPDAVGK
714 568.31 1134.60 1134.58 0.02 0 13 1 YHTEIVFAR
869 412.89 1235.65 1235.70 -0.05 0 (10) 10 LNIPVSQVNPR
871 412.91 1235.72 1235.70 0.02 0 (21) 1 LNIPVSQVNPR
872 618.87 1235.72 1235.70 0.02
909 631.35 1260.68 1260.67 0.01
1093 461.26 1380.76 1380.74 0.01
1717 610.65 1828.94 1828.92 0.02
2191 827.43 2479.25 2479.19 0.06
Proteins matching the same set of peptides:
gi|21361181 Mass: 112824 Tot
ATPase, Na+/K+ transporting, alpha 1 polypeptid
45. gi|1362855 Mass: 99736
lrp protein - human
Check to include this hit in error tolerant sear
Query Observed Mr(expt) Mr(calc) Delta
33 355.72 709.43 709.42 0.01
214 417.24 832.47 832.46 0.01
285 441.27 880.52 880.51 0.00
456 496.78 991.55 991.53 0.01
1276 497.95 1490.82 1490.73 0.09
1701 606.00 1814.98 1814.95 0.03
1756 624.71 1871.11 1871.09 0.02
1906 670.02 2007.03 2007.01 0.02 0 16 1 ELLELEALSMAVESTGTAK + Oxidation (M)
2025 718.04 2151.11 2151.09 0.02 0 26 1 IPPYHYIHVLDQNSNVSR
2026 538.79 2151.12 2151.09 0.03 0 (20) 1 IPPYHYIHVLDQNSNVSR
Proteins matching the same set of peptides:
gi|19913410 Mass: 99266 Total score: 177 Peptides matched: 10
major vault protein [Homo sapiens]
46. gi|11559929 Mass: 97655 Total score: 166 Peptides matched: 8
How do we know if these identifications are correct?
Can't we implement some measure of quality control?
Estimation of false positive rate!

Estimating false positive rates via random databases



Beausoleil, Nature Biotechnol, 2006

Estimating false positive rates in protein identifications

Estimate rate of false identifications by searching a randomized databased composed of forward and reversed protein sequences

```
TEHGK
YDGPLQAK   Forward
LPMVGIR
*KGHET
*KAQLPGDY  Reverse
*RIGVMPL
```

$$\text{FP rate} = \frac{2 \times n(\text{rev})}{n(\text{rev}) + n(\text{forw})} \times 100\%$$

n(forw) = 500
n(rev) = 10

$\longrightarrow$ FP rate = 4%

Protein Identification

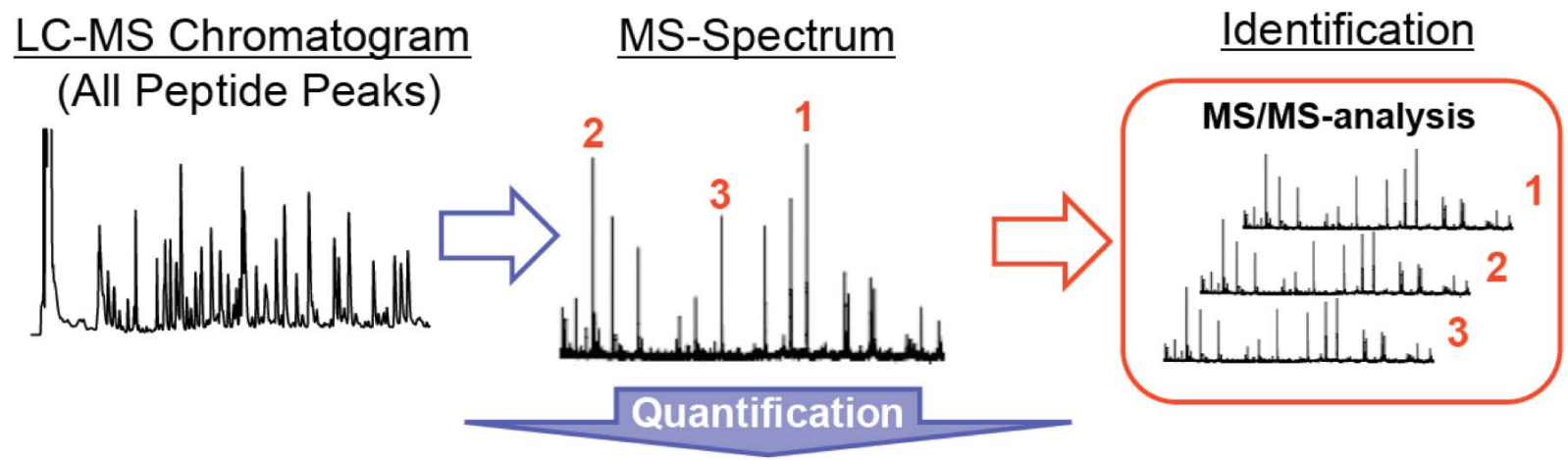
- How can we identify a protein by MS?
- Why do we need to digest protein?
- How can we identify 1000 proteins by MS?
- How do we know they are correct?
- How do we quantify identified protein by MS?

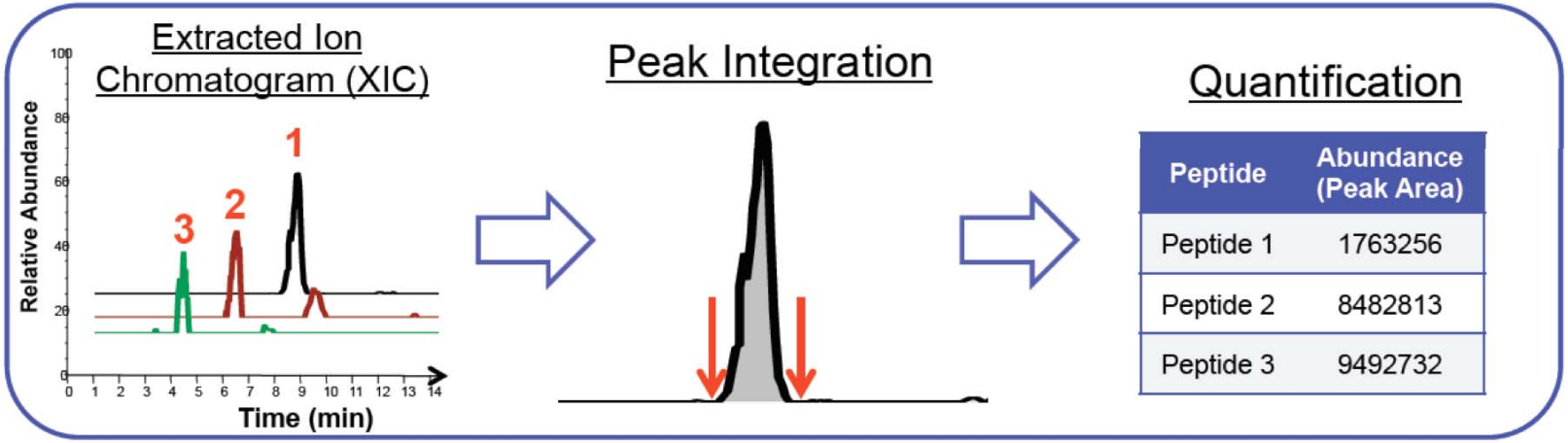
Quantitative Proteomics

Quantification Methods

- Label-free Quantification
  - Automated peak alignment and integration
- Stable Isotope Dilution Quantification
  - General workflow
  - Principles of the popular *SILAC*, *iTRAQ* and *AQUA* methods

Quantitative Proteomics Workflow





Peptide	Abundance (Peak Area)
Peptide 1	1763256
Peptide 2	8482813
Peptide 3	9492732

Quantitative Proteomics

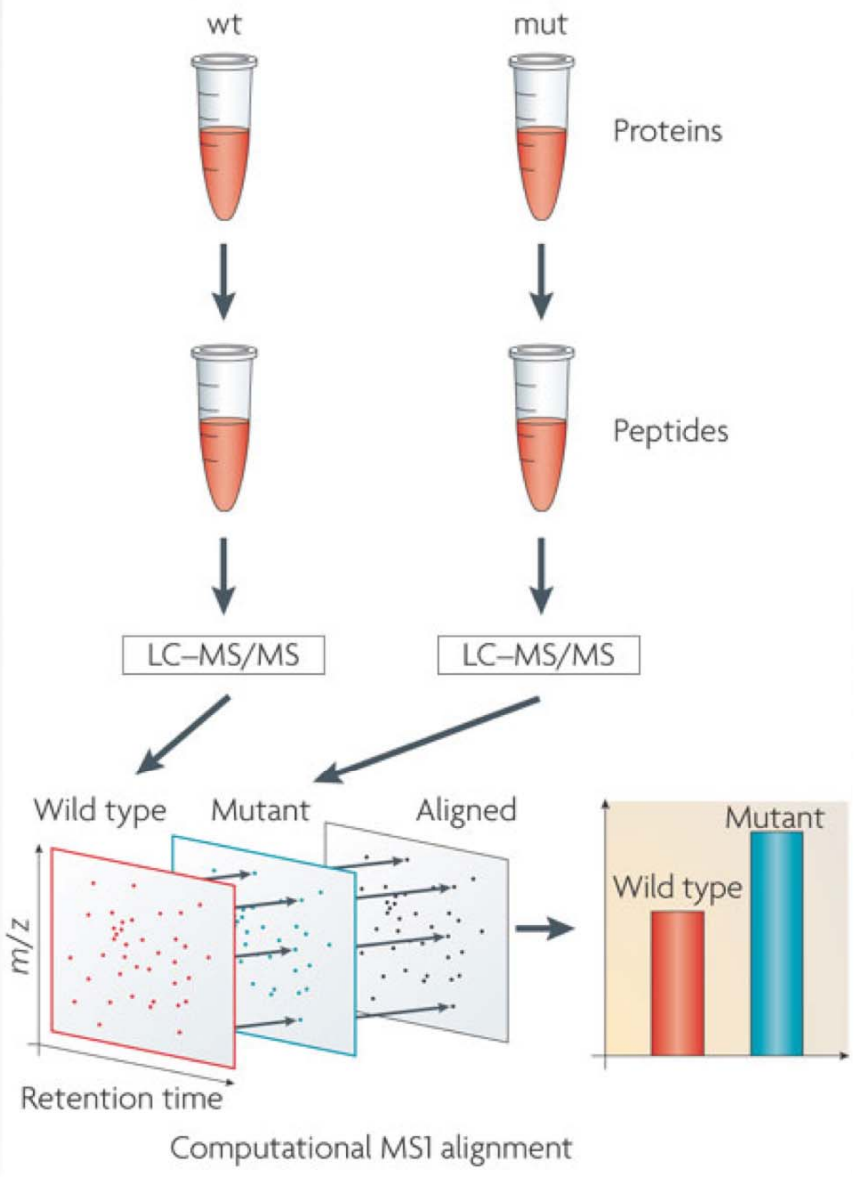
Relative Quantification

- Measure relative intensity of signals of peptides and compare between samples
- This can be employed to determine quantitative differences of **the same peptide/protein** between samples

Absolute Quantification

- Each peptide has a characteristic response factor (F)
- F x relative intensity = peptide concentration
- Usually, a isotopically labeled peptides (internal standard) that is spike it into the sample is used
- This can be used to determine quantitative differences of **different peptides/proteins** within the same or different sample(s)

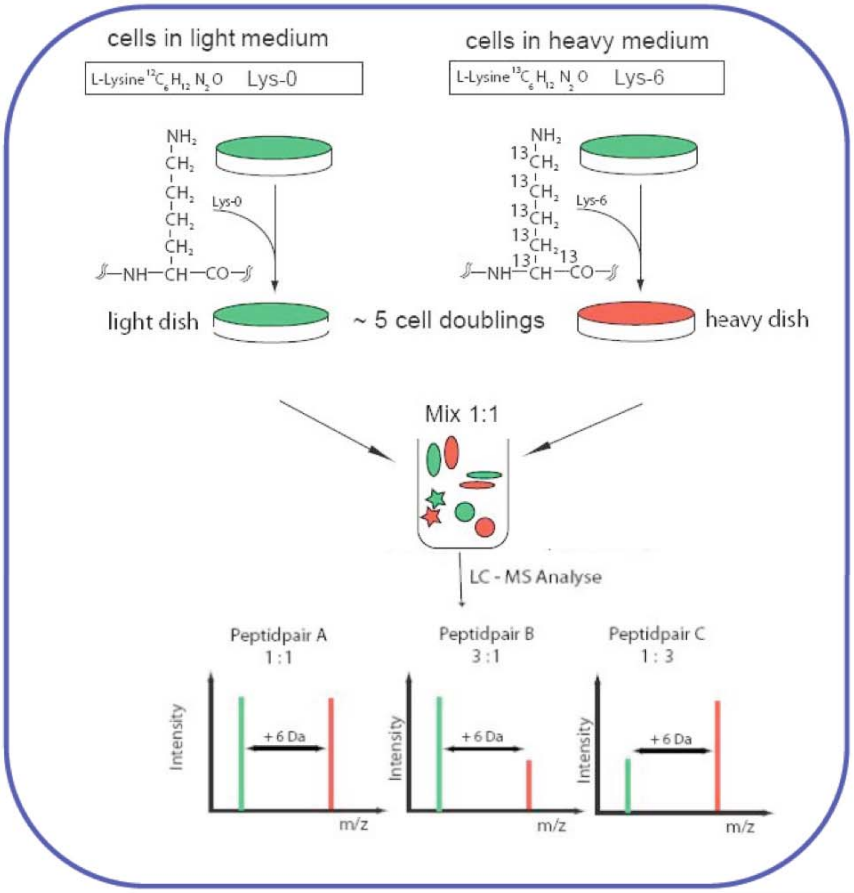
Label-free Quantification



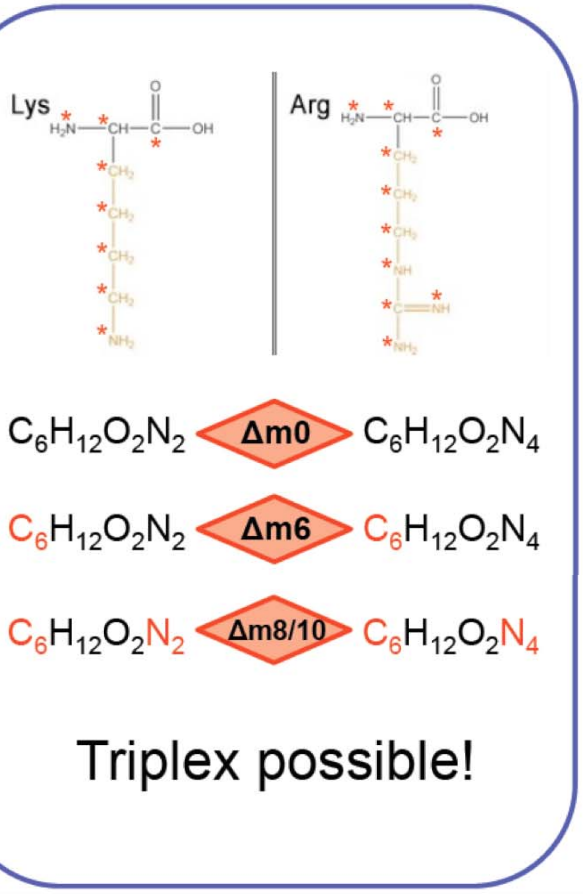
Aebersold R. et al.,
Nature reviews Genetics 2009

Stable Isotope Labeling with AAs in cell culture

SILAC Workflow

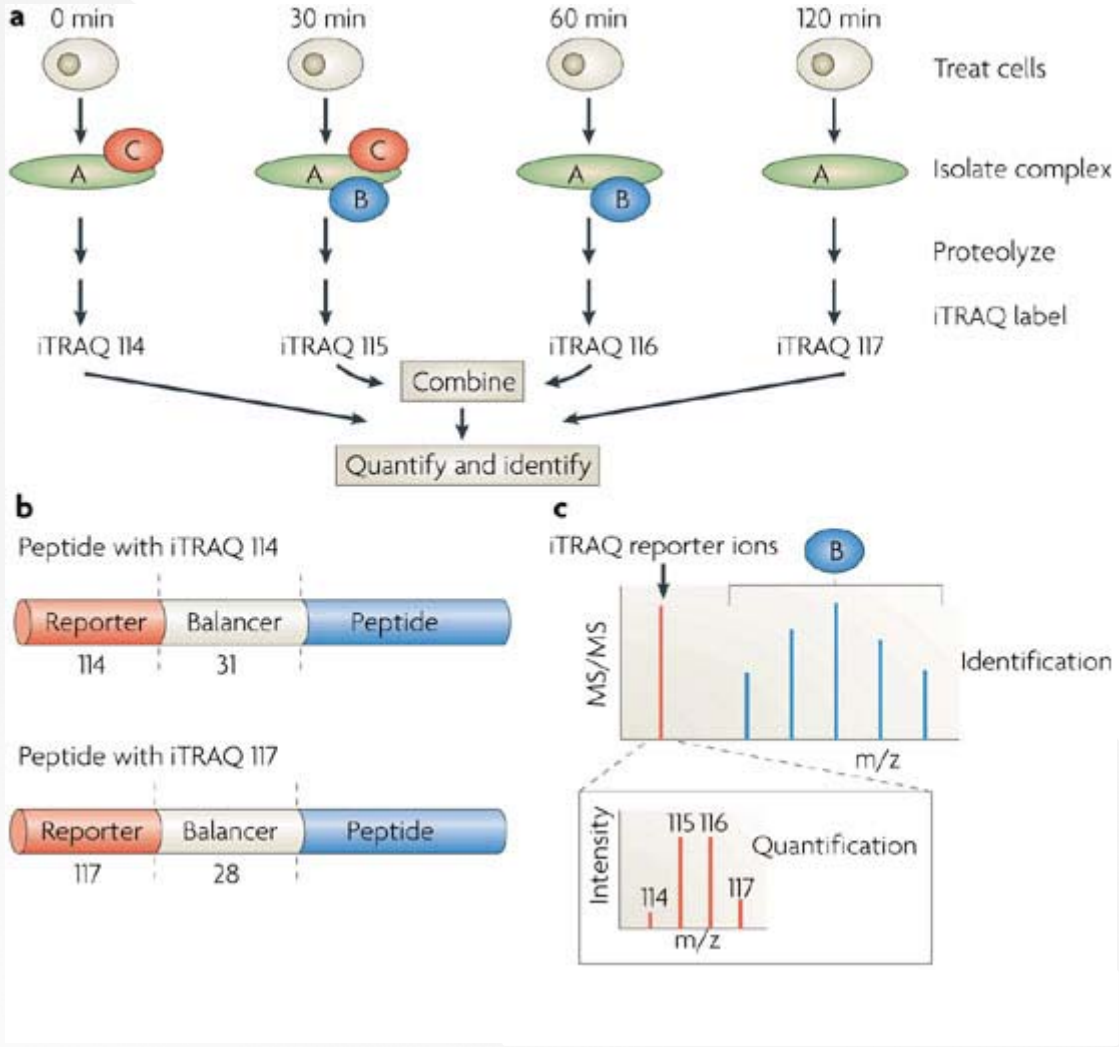


Most suited AAs



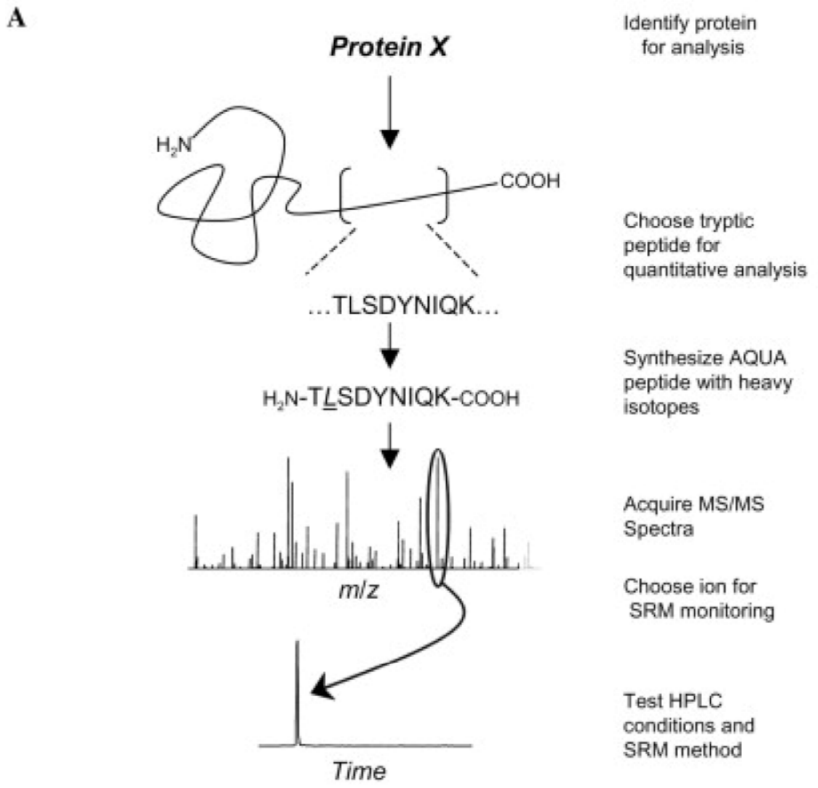
Ong SE. et al., MCP 2002

isobaric Tag for Relative and Absolute Quantitation

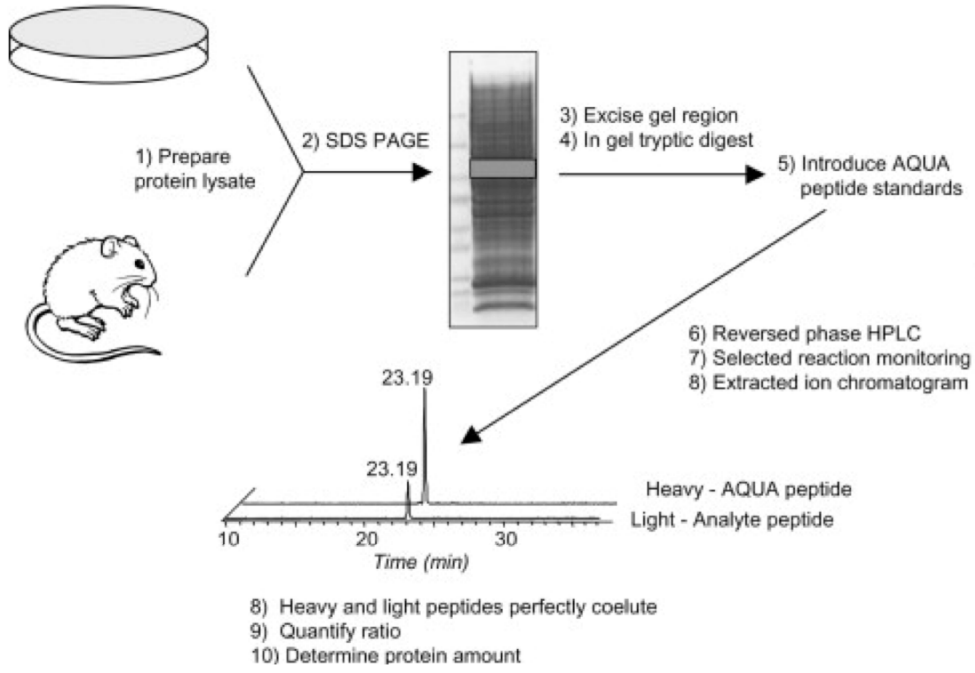


Aebersold R. et al., Nature reviews 2007

Absolute Quantification of Proteins



Stage 1



Stage 2

Gygi SP. et al., Mass Spec. Prot 2005

Take home message

- Always think what you want to do and what do you looking for
- A complex sample need a complex solution
- There is not a uniform method

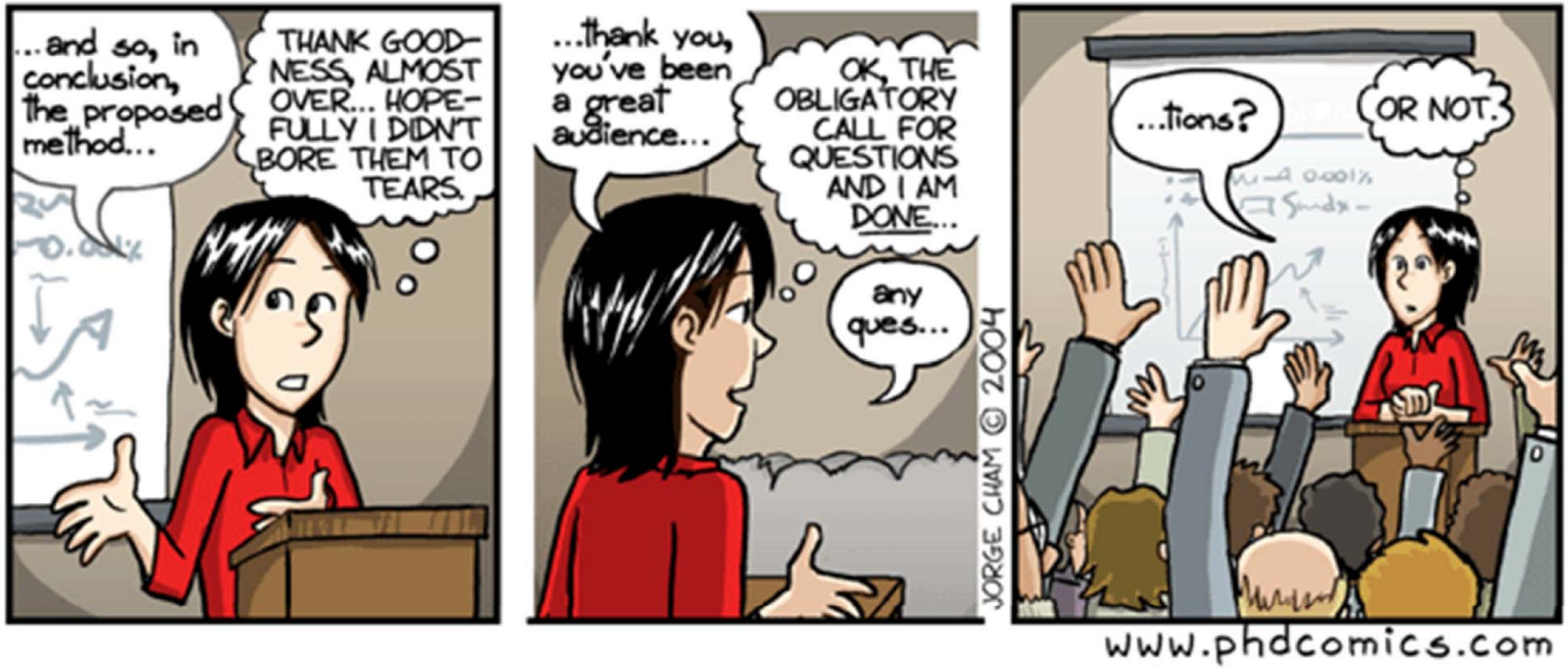
Acknowledgements

To the leader of the MS group

Dr. Cvacka

and members

Thank you for your attention!



any questions?